BBC RD 1991/4

BBC

Research Department Report

IMAGE SCANNING USING A FRACTAL CURVE

J.O. Drewery, M.A., Ph.D., C.Eng., M.I.E.E.

Research Department, Engineering Division
THE BRITISH BROADCASTING CORPORATION

IMAGE SCANNING USING A FRACTAL CURVE

J.O. Drewery, M.A., Ph.D., C.Eng., M.I.E.E.

Summary

This Report describes a novel method of scanning and displaying images based on a self-similar fractal curve instead of the traditional raster. The Hilbert curves are introduced as a subset of the Peano curves and proposed as a means of doing this so as to provide a hierarchically compatible way of evolving from lower to higher definition systems. After describing the properties of the scan and the way in which the hierarchical behaviour is obtained some disadvantages of a practical system are described. The introduction of interlacing and extension to three-dimensional scanning are then treated. Finally, practical experiments are described, designed to test the hierarchical properties, noise performance and susceptibility to timing errors. The conclusions are reached that the assumed hierarchical compatibility works and at the very least the idea could be used as a transmission technique to optimise the use of bandwidth. As a scanning technique, it would be feasible once the technology of discrete source and display devices becomes sufficiently advanced, but would have to be introduced in the context of a completely new service.

Index terms: *Scanning; standards; compatibility; simulation; HDTV; image processing; bandwidth compression; resolution; TV displays; TV systems.*

Issued under the Authority of

Ian Childs

Research Department, Engineering Division
BRITISH BROADCASTING CORPORATION

Head of Research Department

(EL-199)

1991

IMAGE SCANNING USING A FRACTAL CURVE

J.O. Drewery, M.A., Ph.D., C.Eng., M.I.E.E.

IMAGE SCANNING USING A FRACTAL CURVE

J.O. Drewery, M.A., Ph.D., C.Eng., M.I.E.E.

1. INTRODUCTION

Transmission of images by television is conventionally performed by raster scanning in which a signal is formed by scanning the image in a series of parallel lines, either horizontal or vertical. In broadcast television the scan starts at the top left hand corner and finishes at the bottom right hand corner, the scan lines being approximately horizontal. Moreover, each line is scanned in the same direction so that the scan must return to the beginning of each line in, ideally, an infinitesimally short time known as the flyback.

A disadvantage of raster-based scanning is that it is committed to a standard in the form of the number of lines in the picture. This means that once an image scanning standard is chosen the display must adopt the same standard or convert the incoming signal to its own line standard; this is a complex and costly process. Moreover, any curtailment in the bandwidth of an analogue channel used to carry the signal will affect only the horizontal resolution of the image, not the vertical. Excessive curtailment will therefore give an unacceptable image quality, inferior to that which could be obtained with equal horizontal and vertical resolutions.

A further disadvantage of raster-based scanning lies in the flyback. With conventional CRT displays the overhead of scan flyback, with its large rate of change, is often turned to advantage to provide the high voltage needed. However, if the line rate is increased to provide for higher definition formats the flyback time of conventional deflection circuits may become a significant proportion of the line period resulting in an inefficient use of time. For this reason, a boustrophedonic scanning format has been proposed for higher definition television[1] in which successive lines are scanned in opposite directions. This requires a triangular form for the horizontal scan waveform as shown in Fig. 1(a). Although the horizontal flyback is eliminated the sawtooth of the vertical scan must now be modified to a stepped form as shown in Fig. 1(b) to ensure that adjacent lines are parallel. It also requires that non-linearities in the horizontal scan waveform result in a symmetrical function as shown in Fig. 2 to avoid displacing information between consecutive lines. Although this proposal overcomes the flyback problem it still suffers from the earlier mentioned drawbacks of conventional raster scanning.

Until comparatively recently, the scanning in conventional imaging and display devices has been continuous, by way of a continuously moving electronic beam which is deflected by electric, or more usually, magnetic fields. Recently, however, other imaging and display devices have begun to appear which are based on discrete rather than continuous scanning. Thus, imagers based on charge-coupled devices (CCDs) are now well-established. These work by sensing image brightness at discrete locations, the brightness values being read out in a sequence determined by the way in which the device is fabricated. Other devices, based on discrete sensing, exist where the sequence of values can be determined by horizontal and vertical addressing. Similarly, display devices exist wherein the locational sequence of displayed values can be determined by such addressing. The constraints of electron beam raster scanning are clearly inappropriate for such devices although they are, at present, raster scanned for the sake of compatibility with the conventional system. Thus the vehicles exist for adapting the conventional scanning format.

The considerations of scanning standard and signal bandwidth are of considerable importance today when the world is trying to establish the parameters of a new high definition system. Once the new system is established it will have to be in existence for a considerable time to justify the effort expended in

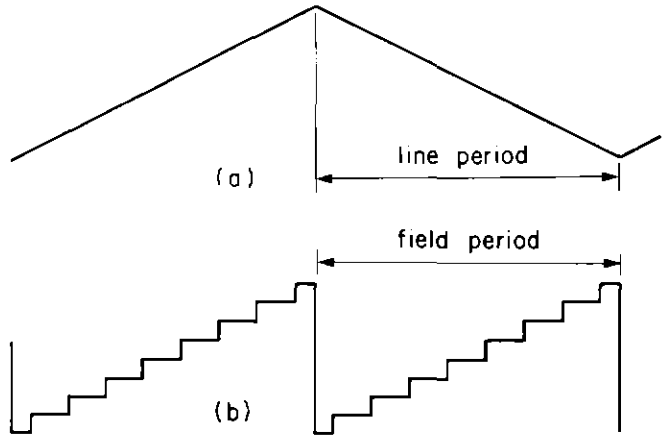


Fig. 1 - The waveforms required for boustrophedonic scanning.
(a) Horizontal (b) Vertical

Fig. 2 - Permissible non-linearity of the horizontal scan.

establishing it and to allow the technology surrounding it to mature. Thus, progress in this field proceeds by a series of jumps. On the other hand, it could be argued that it would be far better if a continual evolution could occur, unfettered by the need to establish scanning standards. This Report describes an investigation into an alternative method of scanning which could form the basis of a system which could offer such a possibility.

2. THE FRACTAL CURVE

It is possible to have a scanning format which is not committed to a line standard and therefore is open to evolution. Such a format relies on the concept of self-similar *fractals*, a term coined by Benoit Mandelbrot. The use of fractals in the field of computer graphics to simulate texture is well established. Moreover, the use of a fractal curve to explore a multi-dimensional space is a known technique[2]. However, the use of a self-similar fractal curve to scan an image for transmission and display, with its attendant advantages, is thought to be novel[3].

It is difficult to give a succinct definition of a fractal without resorting to mathematical symbolism. However, it can roughly be described as a curve or surface whose extent depends on the accuracy of the ruler, i.e. the size of the step used to measure it. This is unlike a conventional curve whose extent tends to a limit as the ruler becomes more accurate. An oft-quoted example is the coastline of Britain whose length increases without limit as the size of the step decreases. Empirically, the number of ruler steps in the length is found to be proportional, not to the reciprocal of the step size, but to the reciprocal of the step size raised to the power of D where D is greater than 1. Thus the length increases as the step size to the power of $1-D$. It is D that Mandelbrot called the "fractional dimension" or "fractal dimension" for short. For the coastline of Britain D is approximately 1.5. A self-similar fractal is one which appears to have the same shape as the ruler size is decreased, i.e. as the magnification is increased. A fuller description of fractals can be found in Reference 4.

The particular fractal chosen for this exercise is the Hilbert curve, which is an example from a set of curves derived by the Italian mathematician Giuseppe Peano[5]. However, other forms of fractal may be more suitable[6]. Although the curve was originally derived by Peano in its two-dimensional form, it may be generalised to any number of dimensions. In particular the three-dimensional form is also relevant in this context although there are certain practical limitations in using it as will be seen.

Fig. 3 shows an example of a Hilbert curve, in this case the fifth-order case joining 32×32 points. Close inspection shows that it can be divided into quarters, each of which is connected by a single step. The quarters, in turn, can be further subdivided into quarters, again connected by only a single step, and so on. Further, the quarters are seen to be replicated, either directly or rotated through a right angle.

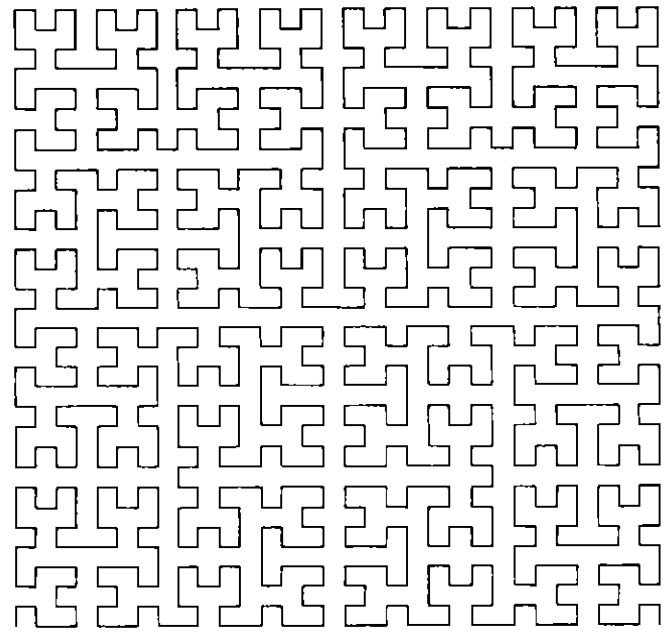


Fig. 3 - A fifth-order Hilbert curve.

Fig. 4 shows the first-, second-, third- and fourth-order curves, each joining double the number of points of the previous order. From these it can be seen that the Hilbert curves form a hierarchical family wherein each curve is derived from that of the previous order. The derivation of one order from the previous one can be carried out indefinitely to yield an abstraction of infinite order which is the fractal curve underlying the family whose members are approximations to it.

It is easy to derive the fractal dimension of this curve. A curve of a particular order, n, joins $2^n \times 2^n$ points. Thus the total number of points is 2^{2n} and the number of steps, N, is given by

$$N = 2^{2n} - 1$$

The step size, s, on the other hand, is given by

$$s = a/2^n$$

where a is the side of the square. Thus

$$2^n = a/s$$

and N is approximately given by

$$N = a^2/s^2$$

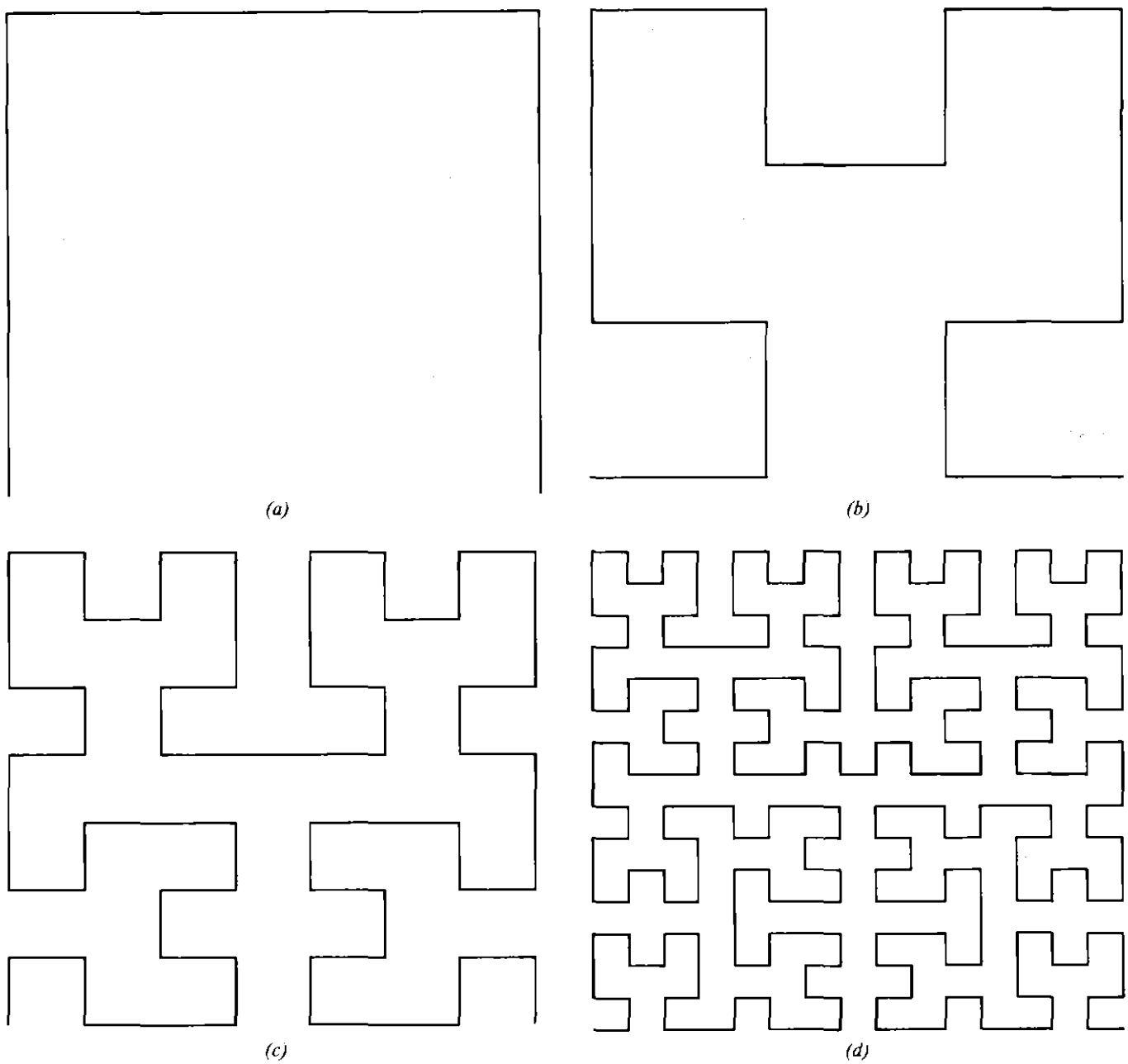


Fig. 4 - The first-, second-, third- and fourth-order Hilbert curves.

So the number of steps, N, is inversely proportional to the square of the step size making the fractal dimension, D, equal to 2. The total length of the curve is in this case inversely proportional to the step size, i.e. it is proportional to the number of steps along each side.

3. THE BASIC IDEA

Imagine, now, that an image is repeatedly scanned with a particular member of the family of curves to form a video signal. If this signal is fed to a display which scans in sympathy, i.e. with the same member of the family in the same phase as the source, then clearly a correct interpretation of the image will be produced. However, what is less clear is that if the display uses any other member of the family then, provided it takes the same overall time to scan the image, a recognisable representation will still be produced because *corresponding parts of the image are scanned at corresponding times.* This gives a means of compatibility between sources and displays. If the display has a lower order than the source this will be defined as *downwards compatible*; if the display has a higher order than the source this will be defined as *upwards compatible.*

This compatibility can be appreciated by considering adjacent orders of scan. Fig. 5 shows the superposition of third- and fourth-order scans. As can be seen, the higher order scan is never more than half

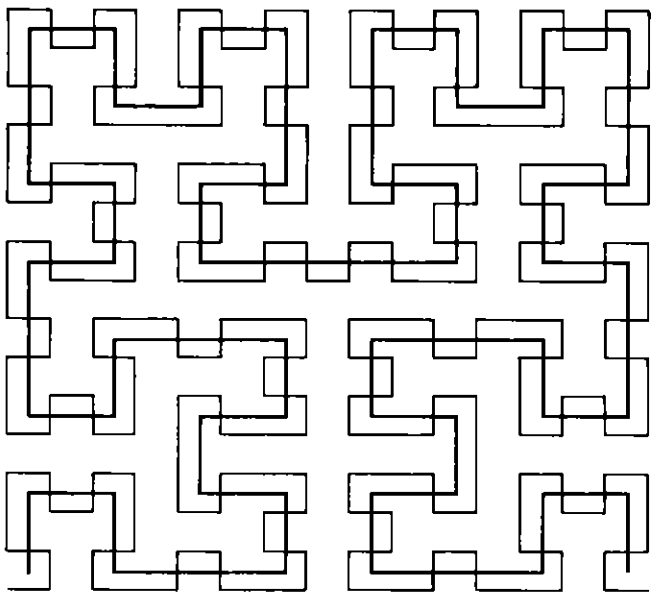


Fig. 5 - The superposition of third- and fourth-order scans.

a step displaced from the lower order. If the scan points represent pixels then four adjacent pixel values of a signal derived on the higher order, if displayed on the lower order, will suffer no more than half a pixel displacement and so will tend to superpose. Equally a pixel value of a signal derived on the lower order, if displayed on the higher order, will be spread about by no more than half a pixel. The mechanism of compatibility will be affected by whether the scanning is discrete, i.e. jumping from point to point or continuous, and this is examined in more detail in Appendix 1.

This compatibility between different situations implies that *it is not necessary to know the order of the scan in order to reproduce a version of the image.* Thus, subject to adequate bandwidth, we have an image transmission system which is independent of the spatial resolution of the source and display provided that they are scanned in this way. Only the number of fields per second, the form of the fractal (i.e. that it is a Hilbert curve) and the starting and finishing points need to be standardised.

An apparent theoretical objection to the idea is that it can be applied only to square arrays of points. Arbitrary image aspect ratios are not possible. However, in the case of Fig. 6, ratios of 2:1, 3:1, 4:1, 3:2 and 4:3 are possible by starting and finishing at the points shown on the figure. Small deviations from these values are possible by altering the pixel aspect ratio without departing significantly from the spirit of isotropic resolution.

Because the scan is contiguous there is no flyback except between fields and thus time is saved. Even field flyback may be eliminated by using a modified square scan consisting of two 2:1 scans joined at their endpoints. In such a case, synchronising information would have to be sparse, compared with conventional scanning, possibly consisting of a recognisable framing pattern.

The implication of the idea is shown in Fig. 7. Cameras operating with various orders of scan may be selected at will and their signals transmitted through the channel. The ninth-order scan camera represents, approximately, the current system. The eighth-order scan camera represents, say, a low-resolution light-weight 'electronic journalism' camera whilst the tenth-order scan camera represents a high definition camera. They could be used in different programmes or even in the same programme. The 11th-order scan camera represents a camera which could be developed in the future, being introduced when required. The same remarks apply to the hierarchy of displays, the eighth-order display representing, say, a wrist-watch receiver based on a liquid crystal display.

Such a concept would, in practice, probably be introduced piecemeal and so it would have to coexist with current scanning technology. If it were introduced on the channel then compatibility with current technology would require raster-to-Hilbert converters in conventional cameras and Hilbert-to-raster converters in conventional displays as shown in Fig. 8, the converters merely re-ordering the sequence of incoming pixel values. Such converters would consist simply of a picture store which would write the pixel data in

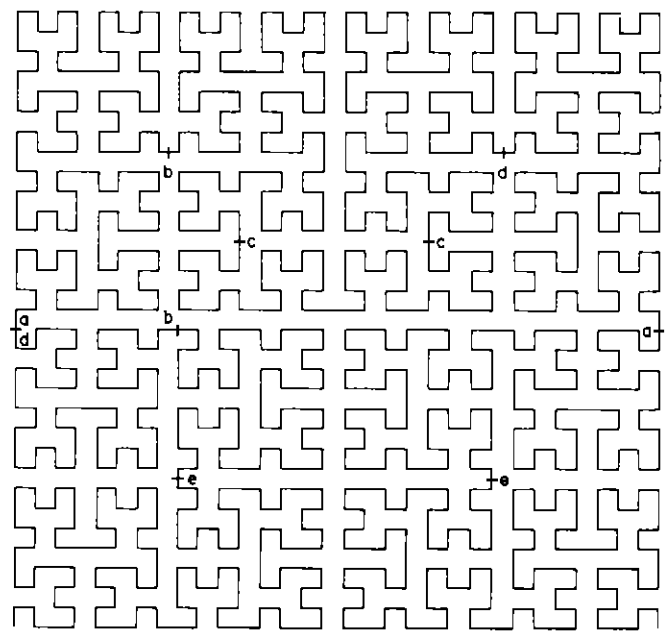


Fig. 6 - Starting and finishing points for various aspect ratios.

(a) 2:1 (b) 3:1 (c) 4:1 (d) 3:2 (e) 4:3

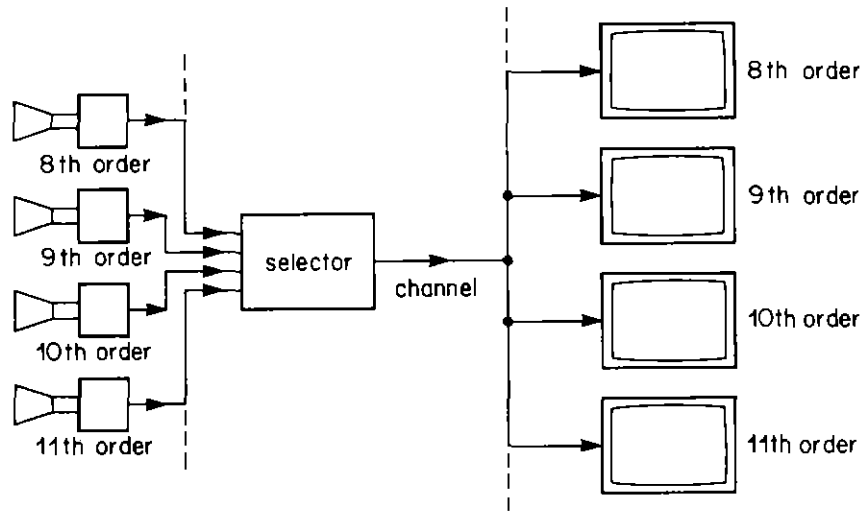


Fig. 7
The basic purpose of the proposal.

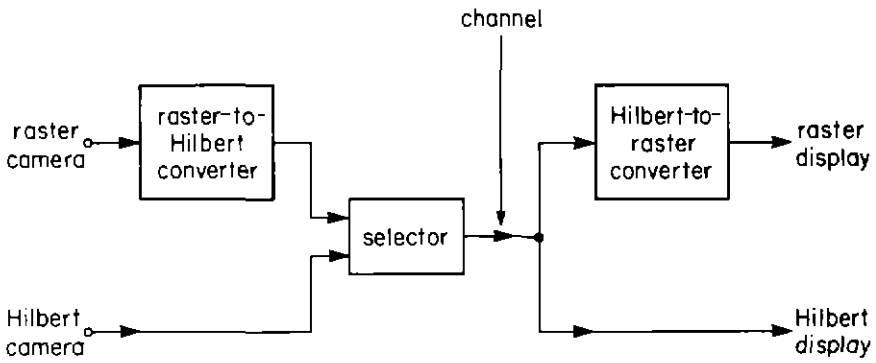


Fig. 8
Conversion at both ends of the Hilbert link.

one sequence and read it in another. This re-ordering of data could also be regarded as a scrambling technique to prevent unauthorised reception of a newly-introduced service.

4. SIGNAL PROPERTIES

4.1 Scan smoothing

The Hilbert curve may be represented parametrically by the sequences of the x and y coordinates of the points as a function of time. For a continuous scan these sequences may be used to generate the deflection waveforms that would need to be applied to a scanning device. Interpolation would occur through the normal bandwidth limiting of physical devices and the nature of this interpolation affects the shape of the scan.

The simplest assumption is that the samples are interpolated with a filter that cuts at half the repetition frequency of the samples. This ensures that the interpolated waveforms pass through the sample values themselves. Fig. 9 shows the result on the Hilbert curve. Note how there are regions of high curvature even though the individual deflection waveforms are smooth. This result is at first sight surprising but inevitable if the curve has to 'get round' the samples and it means that it cannot be traversed at constant speed.

Now consider what happens when the deflection waveforms are smoothed further. Suppose this smoothing is performed by taking a running average of a number of samples and interpolating these averaged samples with the same filter as before. Figs. 10 - 12 show the effect of taking an average over 2, 3 and 4 samples respectively. It can be seen that the shape of the curve changes gradually towards the next lowest order, passing an intermediate stage where there are significant diagonal components. This effect can be further demonstrated by taking higher averages. Figs. 13 and 14 show the effect of averaging over 8 and 16 samples respectively from which it can be seen that such averaging appears like two- and four-fold averaging but one order lower.

Thus it can be seen that smoothing the deflection waveforms causes the curve to 'gracefully degrade' to a lower order. As a result, a recognisable image will still be produced because of the downwards compatible property derived in Appendix 1.

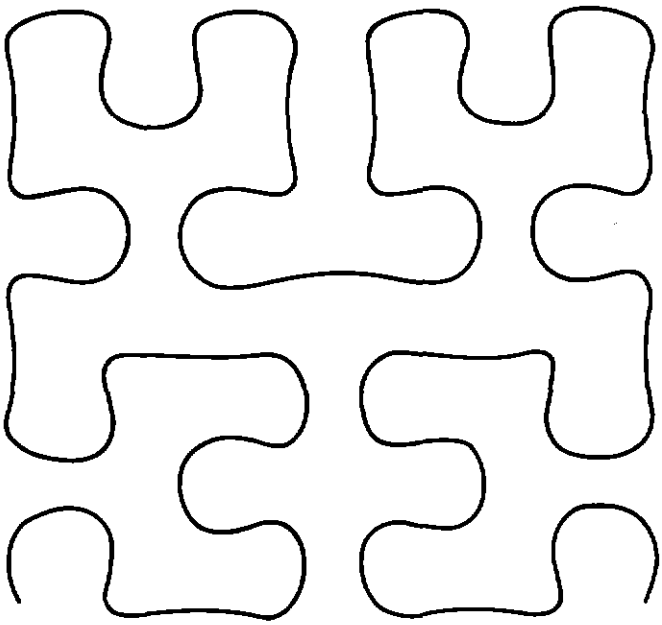


Fig. 9 - The smoothed third-order curve resulting from band-limiting the coordinate waveforms fairly sharply to half the pixel frequency.

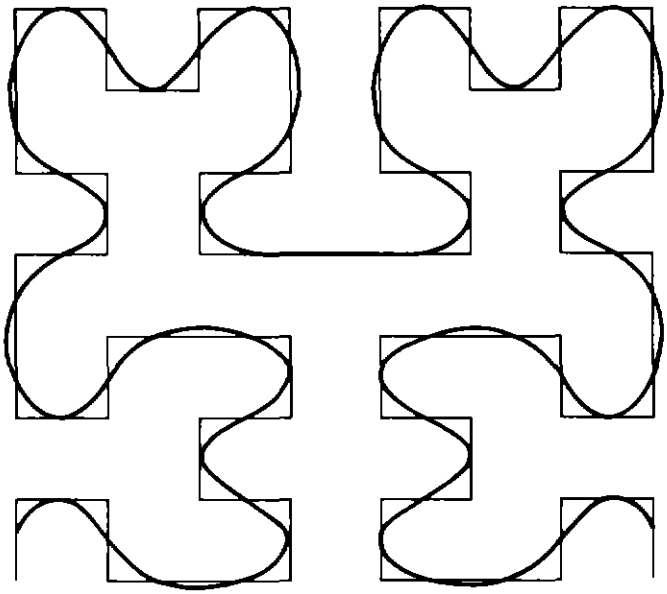


Fig. 10 - Two-fold averaging of the scan.

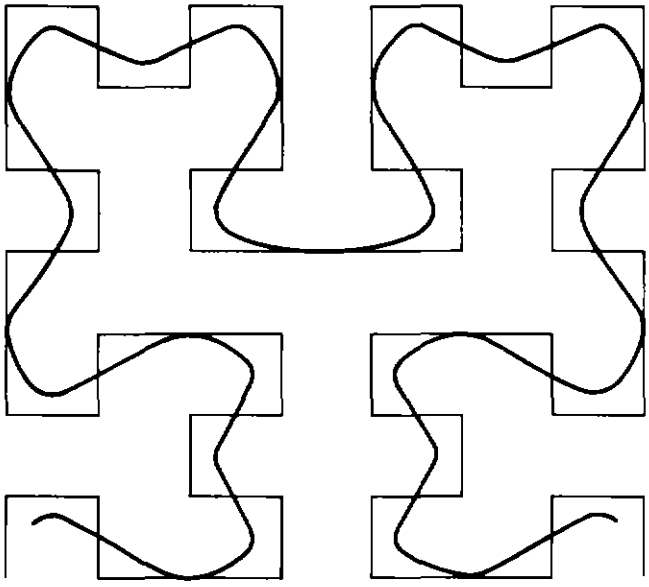


Fig. 11 - Three-fold averaging of the scan.

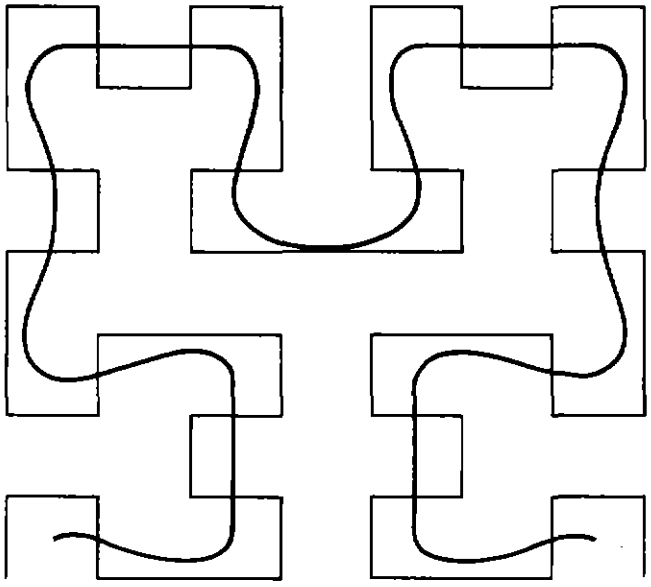


Fig. 12 - Four-fold averaging of the scan.

4.2 Video smoothing

The video signal produced by the source may be considered as a succession of samples or a continuous signal. If it has to be sent through an analogue channel then it must be in continuous form and may be subject to filtering caused by a bandwidth restriction, causing loss of resolution when the signal is displayed. Because the scan never takes more than three steps before changing direction, the smoothing of the video signal applies an equal resolution loss in both directions. Thus we have a system in which bandwidth restriction affects horizontal and vertical resolution equally which is the optimum way of disposing the bandwidth resource. With conventional raster scanning, bandwidth restriction affects only one dimension which is sub-optimum.

4.3 Spectra

One apparent disadvantage of the Hilbert scan is that the scanning waveforms have a bandwidth as

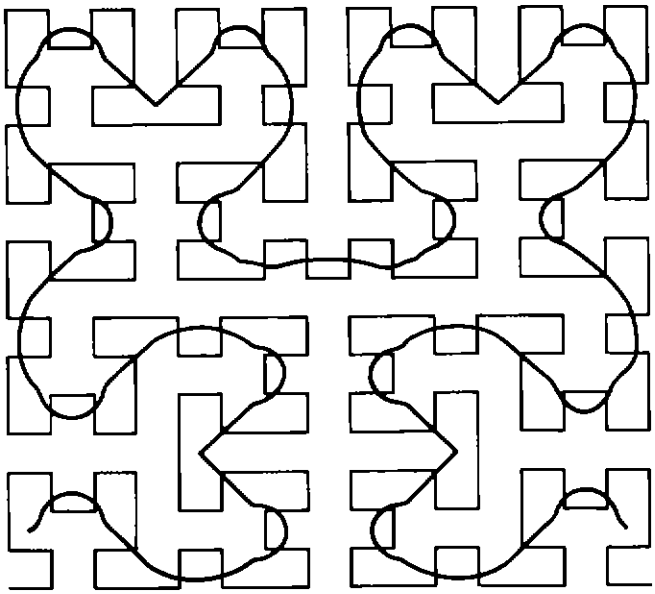


Fig. 13 - Eight-fold averaging of the scan.

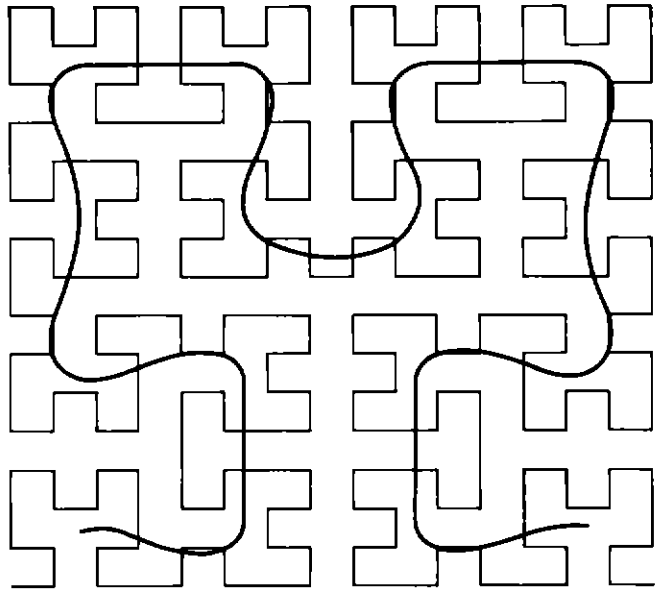


Fig. 14 - Sixteen-fold averaging of the scan.

high as the displayed video signal itself. However, in fairness, it must be pointed out that this is also true of the idealised sawtooth waveforms involved in conventional raster scanning but bandwidth limitation of these waveforms does not upset their primary property of linearity unless it is severe. We have seen that bandwidth limitation of the Hilbert scan causes it to degrade gracefully to a lower order. Thus lack of scan bandwidth is not disastrous but only leads to resolution loss.

This benign behaviour can be explained by the spectrum of the scanning waveform. Inspection of the Hilbert curve shows that there are no abrupt changes of position. Thus we can say that the scanning waveforms do not move very far very fast. In other words the high frequency components they contain are not of high amplitude and the spectrum can be expected to decrease quickly. Fig. 15 shows the waveforms and Fig. 16 shows the spectrum of one of them (the spectra are similar) from which it can be seen that the decrease of amplitude with frequency is dramatic, the bulk of the energy being contained in the first few frequency components. The shape of the spectral distribution is quite strange, having an apparently random characteristic within a gradual decay. It might fancifully be described as having a fractal quality.

Equally, the video signal obtained by Hilbert scanning has similar properties. Figs. 17 and 18 show the spectrum of a portion of an image scanned with a raster and a Hilbert curve respectively. It can be seen that whereas the raster-scanned picture has a spectrum which bunches characteristically about multiples of the line scanning frequency, the Hilbert scanned picture has no such effect. Compared with the raster-scanned picture the spectrum of the Hilbert-scanned picture is 20 - 30 dB down. This suggests that it may be more tolerant to bandwidth restriction.

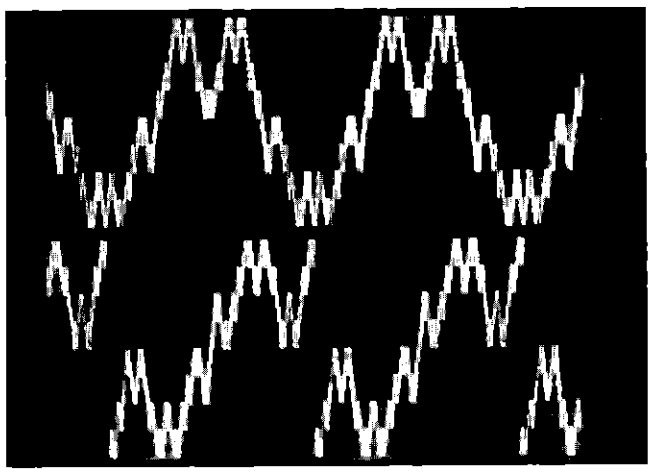


Fig. 15 - The x and y scan waveforms.

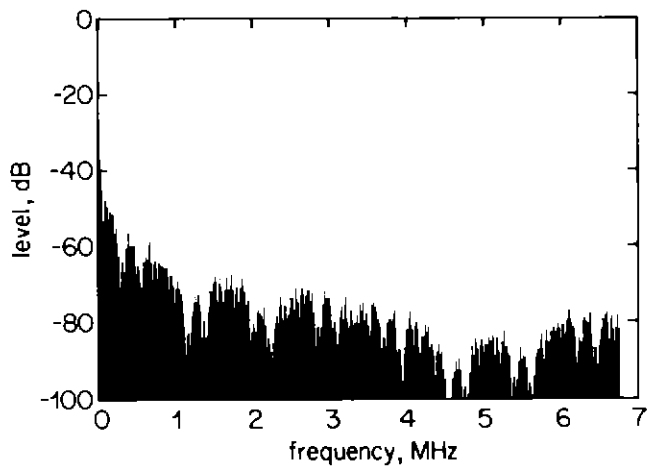


Fig. 16 - The spectrum of the scan waveforms.

This rapid decay of the spectrum implies that the autocorrelation function of the video signal is much broader than that derived by conventional scanning. This fact has long been known and taken advantage of as the basis of data compression systems. A further fact which flows from this is that it should be possible to decrease the channel spacing of transmissions carrying Hilbert scanned signals. This is because adjacent channel interference should be lower due to the quicker spectral decay.

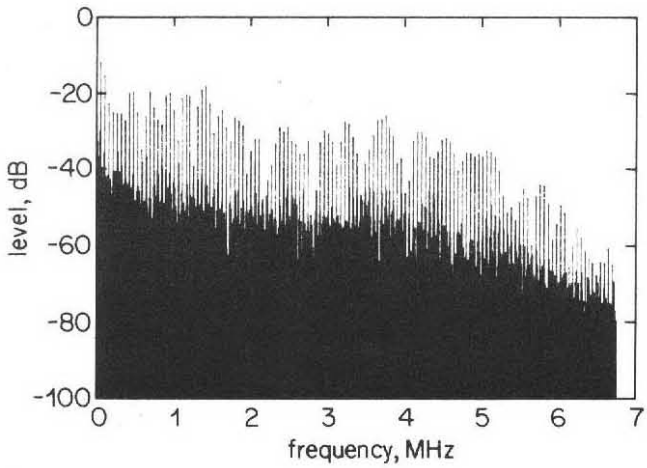


Fig. 17 - The spectrum of the video signal obtained by raster scanning.

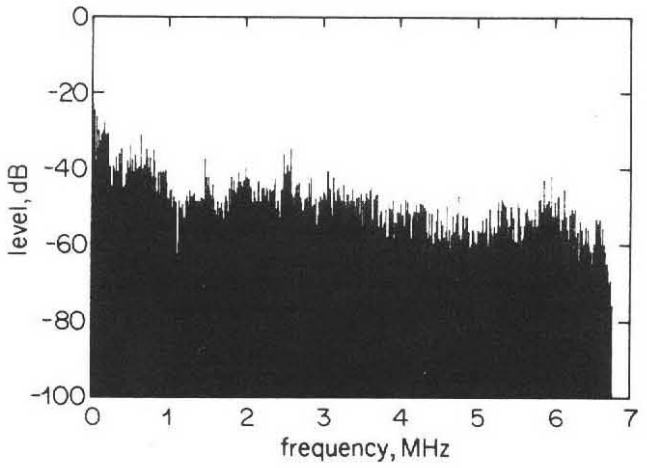


Fig. 18 - The spectrum of the video signal obtained by Hilbert scanning.

4.4 Adjacent channel performance

Fig. 19 shows the appearance of an adjacent channel whose carrier lies one and a half video bandwidths from the prime channel which carries a uniform grey. Raster and Hilbert scanned pictures are shown for comparison under identical conditions. It can be seen that the images are discernable only in areas of fine detail as might be expected. But whereas the raster picture has a vertical line structure in these areas (probably brought about by the precision line-offset chosen) the Hilbert picture has a more mottled appearance. On balance it may be thought that the Hilbert picture is more noticeable. Fig. 20 shows the appearance for a carrier on the upper edge of the video band, i.e. total overlap. Now the raster signal is more noticeable than the Hilbert signal. This tends to suggest that the greatest benefit comes from the behaviour of the low frequencies.

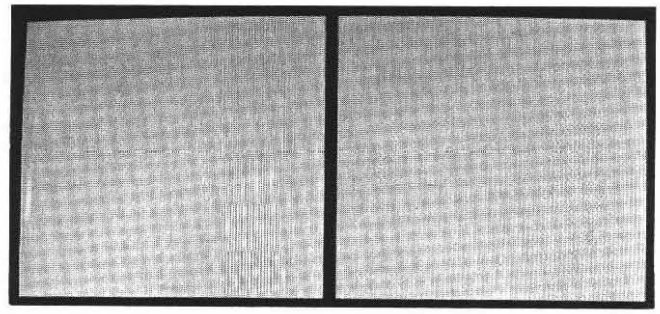


Fig. 19 - The appearance of the adjacent channel for a carrier spacing of one-and-a-half video bandwidths.

Raster scanning *(left)* Hilbert scanning *(right)*

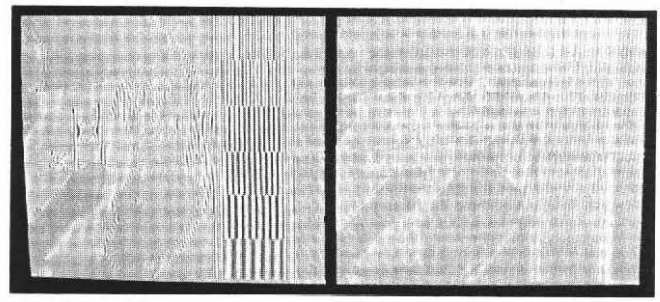


Fig. 20 - The appearance of the adjacent channel for a carrier spacing of the video bandwidth.

Raster scanning *(left)* Hilbert scanning *(right)*

5. PRACTICAL OBJECTIONS

The objections to introducing Hilbert scanning fall into two categories, instrumental and fundamental. Those in the former are mainly concerned with the problems of continuous scanning whilst those in the latter include compatibility with the current system.

5.1 The effects of finite bandwidth on the scanning waveforms

For a continuous scan the curves of Figs. 3 to 5 are idealised in that if the curve is traversed at constant speed, the x and y coordinate waveforms will have discontinuities of gradient and thus possess infinite bandwidth. Any physical device will have a finite bandwidth and thus the coordinate waveforms will be smoothed to some extent. This will, in general, cause the curve to depart from the idealised straight form, as noted in the previous section.

There is, however, a special case where limiting the bandwidth of the waveforms does not smooth the curve. This is when the impulse response of the interpolating filter is a raised cosine of base width equal to twice the point interval. Such a 'raised cosine' filter has a spectral characteristic which is limited, for all practical purposes, to twice the reciprocal of the base width of the pulse, i.e. the point frequency. In other words, the bandwidth is limited to twice the Nyquist limit for the sequence of point values. Fig. 21 shows a portion of the waveforms when so filtered. As can be seen, at the ideal point values, each coordinate has zero rate of change and neither overshoots. This means that the spot comes momentarily to rest and, as only one coordinate changes at any time, the motion passes smoothly from one coordinate to the other. Thus idealised scan geometry implies uneven speed.

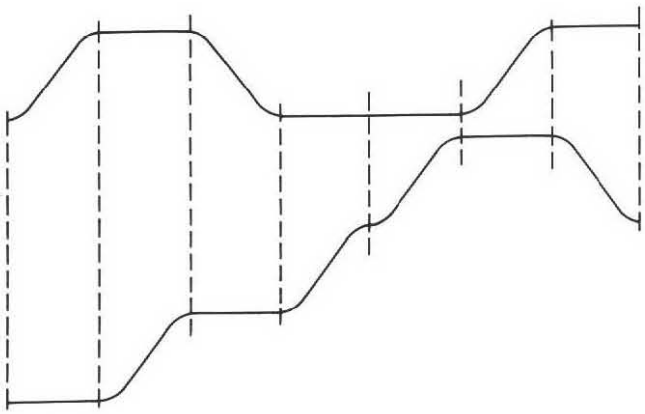


Fig. 21 - A portion of the horizontal and vertical waveforms interpolated by a raised cosine filter.

5.2 Speed modulation

This complication of uneven speed makes analysis difficult. In the imaging device, the video signal will be modulated by the speed, being greatest where the speed is least and vice versa. In the display, the light output will be greatest where the speed is least and vice versa. This effect cannot be avoided, only minimised. Moreover, the situation becomes even more difficult to analyse if the finite size of the beam is taken into account. For then modulation will occur even if the speed is constant, by virtue of the change in direction at most points. For example, Fig. 22 shows the variation in brightness of the display trace at a corner traversed at constant speed by an idealised square aperture. Note the discontinuities in brightness in spite of the constant speed. Fig. 23 shows the variation at three successive corners (typical of most of the scan) when the aperture size equals the point spacing. It can be seen that in this example one of the discontinuities is eliminated. For this reason it is of interest actually to perform the image display with the scan to discover how important these factors are.

5.3 Electron optics

Given that the required bandwidth of the scan may very well be higher than that of the video, it may then be argued that it would be impossible to deflect a beam with such a bandwidth using magnetic fields, only electrostatically. This argument could be based on the magnetic 'inertia' of the deflection system caused by the inductance of the deflection coils. Certainly it is known that calligraphic deflection displays can suffer from magnetic hysteresis but this is usually in the context of large positional swings. However, the rate of change of position in a Hilbert scan is actually less than that of the flyback in a conventional scan so that the objection cannot be sustained. Unfortunately, it has not yet been possible to test the hypothesis with a calligraphic display using magnetic deflection as these are difficult to procure.

As mentioned already, these objections are irrelevant to discrete sources and displays. If they proved to be well-founded then the video signal obtained from a raster-scanned source could be converted to Hilbert form by merely reordering

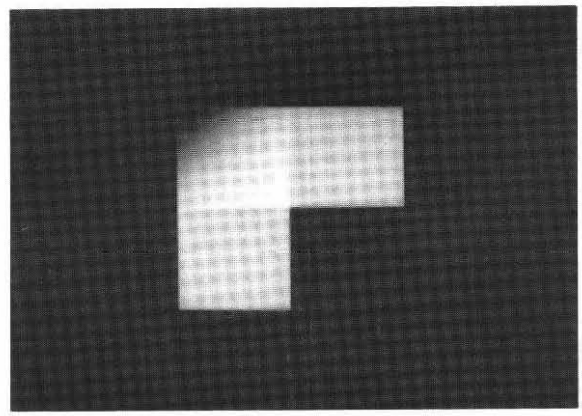


Fig. 22 - The variation in display brightness at a corner traversed by an idealised rectangular aperture at constant speed.

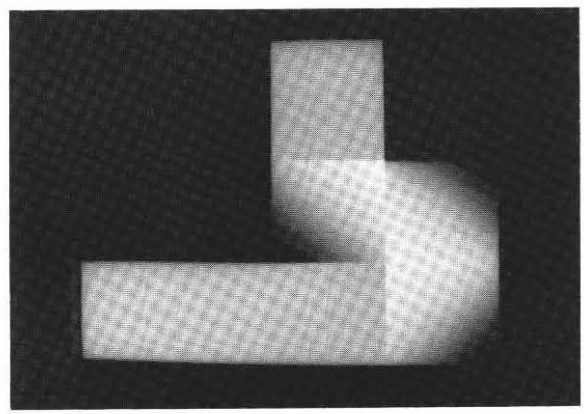


Fig. 23 - The variation in brightness at three successive corners typifying a part of the scan when the aperture size equals the corner spacing.

samples using a picture store. Likewise a raster-scanned display could include a conversion from Hilbert form by a similar mechanism. However, as will be seen, this might cause problems of compatibility with motion.

5.4 Synchronisation

An objection of more substance is that, as the scan has no flyback except between fields, there is no opportunity for establishing black level or synchronisation except at field intervals. Whilst this is true for the pure scan, there is nothing to stop the points at the scan edges representing black level so that the reference and synchronising information may be contained in these regions of the signal. While these regions do not occur at completely regular intervals, their occurrence can be predicted whatever the order of scan. Fig. 24 shows the times at which the edge points occur for the first few orders. As the resolution increases these points become more numerous but, of course, they do not occur while the scan curve is away from the edge of the field of scan. For example, no points occur in the central sixth of the scan.

In this connection, a problem could arise with hierarchical compatibility. If a high resolution signal is being displayed with a low resolution system, then, on the basis of the display resolution, the synchronising information would appear to be interspersed with picture material and so be diluted. Further consideration is necessary here. If, on the other hand, a low resolution signal is being displayed with a high resolution system no problem arises as the synchronisation information would now appear to occupy picture material as well and so be easier to extract.

5.5 Compatibility

Naturally, the severest objection is that such a scanning method is completely incompatible with the present systems. Thus it could not be introduced 'overnight' into the existing system but only through the medium of a completely new service. If this were done, however, the exercise would never need to be repeated however often the resolution of the service were improved. Even with the present systems, the method offers a coding strategy for transmission using pixel rearrangement at each end of the channel wherein its hierarchical properties allow it to make optimal use of transmission bandwidth.

6. INTERLACE

Aside from instrumental considerations, a fundamental disadvantage of Hilbert scanning as so far described, compared with conventional raster-based scanning, lies in the bandwidth needed for transmission of the video signal. As with conventional scanning, a field may be defined as one traverse of the image and, to avoid perceptible flicker, the field frequency must be higher than a limit set by the brightness of the display. This limit is generally agreed to be in the region of 50-60 Hz. Now, to conserve bandwidth, conventional raster scanning transmits the brightness of image points on only alternate lines of the image in a single field, the brightness of the points on the remaining lines being transmitted during the following field. This is termed 'interlaced scanning' and halves the bandwidth compared with a non-interlaced scan whilst preventing flicker by maintaining the refresh frequency of a region in the display. It is effective so long as the vertical detail in the image is not enough to cause the brightness content of spatially adjacent lines to differ markedly. If it does so, then a disturbing flicker at half the field frequency, known as 'inter-line twitter' is produced. Given that this technique is generally effective and used universally for image broadcasting, it can be seen that, unless a similar technique is applied to Hilbert scanning, it will be at a disadvantage as, otherwise, all the brightness values must be transmitted in one field period, giving rise to a video signal having twice the bandwidth of a conventional signal.

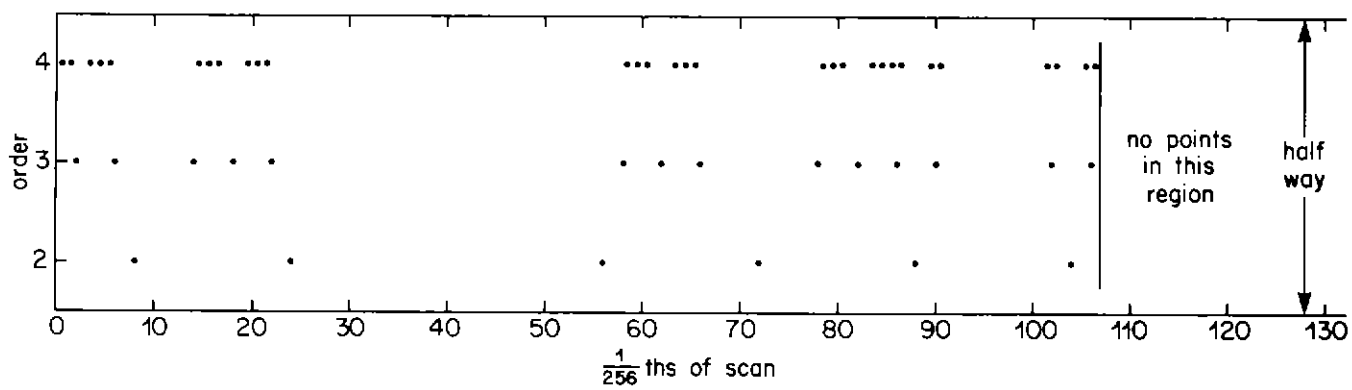


Fig. 24 - The occurrence of edge points for different orders of scan.

Following the analogy with conventional scanning, it may be thought that the concept of interlaced scanning may be extended to the Hilbert scan by interlacing scans of lower order. Fig. 25 shows an attempt to do this by interlacing two second-order scans to form a third-order scan. In practice this fails to work because, as can be seen, some points are not visited whereas others are visited twice. This results from the fact that the curve of the next lowest order connects only a quarter of the points. Moreover, even if this difficulty were overcome by modifying the interlaced curves, the hierarchical advantage would be lost. This is because the spacing of the interlaced curves in the next lowest order would be doubled whereas the information in the video signal would correspond to the original spacing and thus be inappropriate.

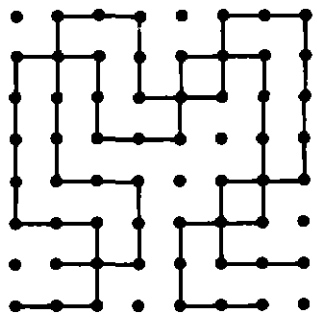


Fig. 25 - An attempt at extending the concept of conventional interlaced scanning by interlacing two lower-order scans.

An alternative approach, which conveys all the information and which preserves the hierarchical property, is to recognise that only a single curve is needed, the interlace being obtained by sample selection *along the curve*[7]. Thus, at the source, the image is scanned with the Hilbert curve but only, say, the odd samples are taken in the first field with the even samples being taken in the second field as shown in Fig. 26. If the source is discrete then this is straightforwardly done through the addressing sequence. If the source is continuous then there are two ways this may be done. In the first method, which is preferred, the scan traces the full Hilbert curve and the signal so obtained is sampled at the pixel rate, alternate samples then being discarded. The remaining samples are then interpolated to provide an analogue signal. In the second method, the scan traces only the appropriate points, the analogue signal being obtained directly. In this case the scan should be such as to spend a reasonable proportion of each sample period at the sample points and a relatively small proportion of the period moving between the points.

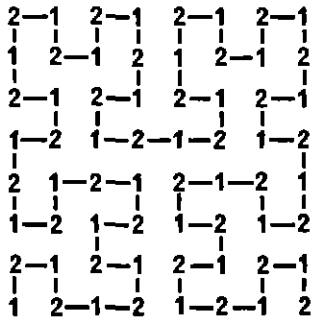
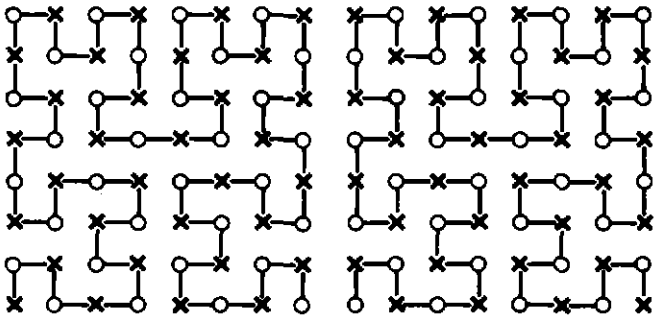


Fig. 26 - Interlaced scanning by sample selection at the source.

At the display, the incoming video is arranged to illuminate only those points corresponding to the sample values that are sent. Again if the display is discrete then this can be done straightforwardly through the addressing sequence. If the display is continuous then again there are two ways this may be done. In the first method, which is preferred, the display scan traces the full Hilbert curve and the signal is augmented with samples corresponding to black-level at, say, the even positions in one field and the odd positions in the other field as shown in Fig. 27. The insertion of the black-level samples on one field creates the spaces into which the interleaved samples on the other field fit. The frequency of insertion must be at half the display pixel rate which is not necessarily the incoming rate. If it is, the operation



(a) *(b)*

Fig. 27 - Interlaced scanning at the display.
(a) Prime field (b) Interlaced field

doubles the bandwidth of the signal but only locally, in the display circuitry. In the second method the display scan traces only the appropriate points and the signal may be used directly without black sample insertion. As in the second method at the source the same qualification applies concerning the movement of the trace.

With either method, some form of ident signal must be present on each field to indicate field parity to the receiver. This is because there is an even

number of points in the scan and so the sample selectors at both ends of the system must be reset at the field boundaries. If there were an odd number of points in the scan, the ident would be unnecessary because alternate sample selection would have a period of two fields and the correct phase of alternation at the receiver would be ensured by the phase of the field synchronising pulses, just as in conventional scanning. However, it is impossible to preserve the hierarchical properties if there is an odd number of samples in the scan and so, reluctantly, this idea must be abandoned.

Just as with conventional scanning, this interlacing will be effective only so long as the image is not fine enough to create a significant difference in the content of the two sets of sample values. With conventional scanning such differences occur on sharp horizontal edges or fine horizontal gratings whereas Fig. 26 shows that, with Hilbert scanning, it occurs on diagonal edges or gratings. Inasmuch as these occur less often and the spacing of the diagonal rows is finer, the situation is less likely to occur. Thus, Hilbert scanning has a significant advantage over conventional scanning in this respect.

It is shown in Appendix 2 that such a system preserves the hierarchical properties of upwards and downwards compatibility in that any source scanning order may be combined with any display scanning order. Moreover, the preferred method of scanning at the display may be used with either interlaced or non-interlaced scanning at the source so enabling an evolution from interlace to sequential scanning to occur at the source.

The method may be generalised to arbitrary interlace factors wherein only every qth point is transmitted on each field, different sets being transmitted successively, taking q fields to transmit a full picture. In such a qth order interlace there are $(q - 1)$ factorial ways of defining the sequence of sets. In such a situation it is useful to define the sequence, where possible, by the expression

$$ip \text{ modulo } q$$

where i is any integer, q is the interlace order, p is about one half of q and p and q are coprime, i.e. share no common factors. For example, if q is 12 and p is 5 then the sequence

0, 5, 10, 3, 8, 1, 6, 11, 4, 9, 2, 7, 0, ...

is obtained. An advantage of using such a high order interlace is that the field frequency may be increased considerably without increasing the bandwidth of the transmitted signal.

7. THREE-DIMENSIONAL SCANNING

As stated earlier, the Hilbert curve is a particular case of a set of curves derived by Peano and is applicable to any number of dimensions. Fig. 28 shows the three-dimensional version of the second-order case which joins a cube of $4 \times 4 \times 4$ points. The curve is built up from elementary sections joining $2 \times 2 \times 2$ points shown in Fig. 29, which can be obtained from Fig. 28 by selecting every eighth running average of eight points. Clearly, then, smoothing the curve causes it to degrade gracefully to the next lowest order just as in the two-dimensional case. Higher orders can be derived joining double the number of points in each dimension.

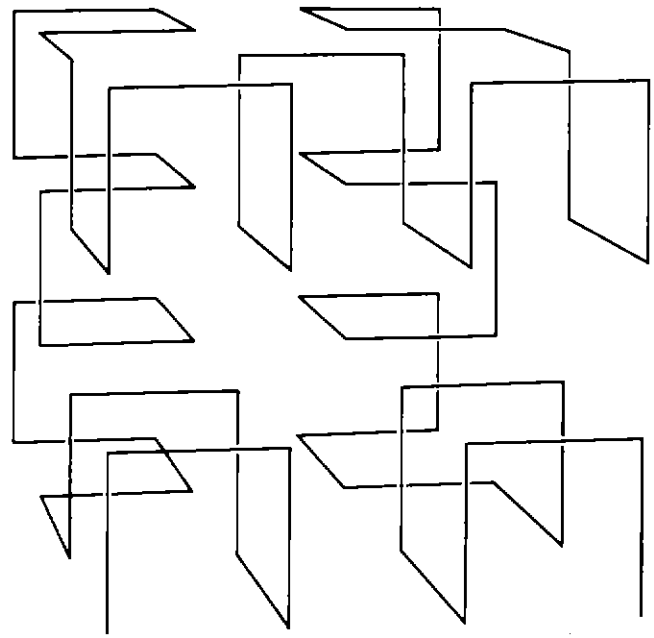


Fig. 28 - The three-dimensional second-order Peano curve.

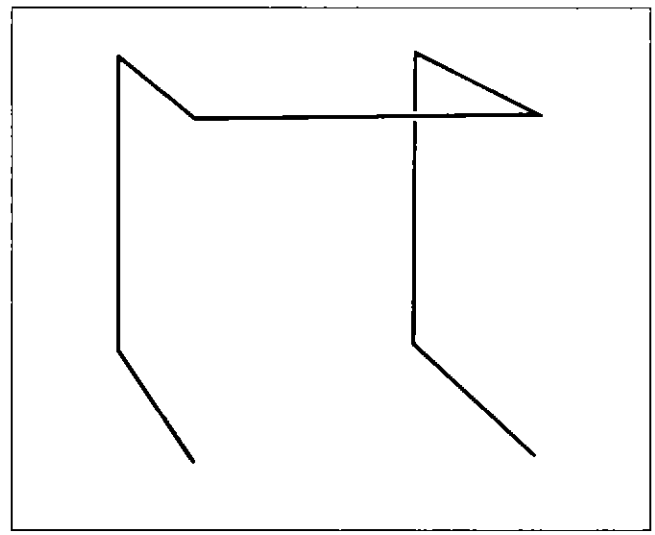


Fig. 29 - The three-dimensional first-order Peano curve.

The fractal dimension of this curve may be derived by an argument similar to that in Section 2. Thus the number of steps, N, in a curve of order n is given by

$$N = 2^{3n} - 1$$

which is approximately

$$N = a^3/s^3$$

where a is the side of the cube. Thus the number of steps is inversely proportional to the cube of the step size making the fractal dimension equal to 3. The length of the complete curve is now inversely proportional to the square of the step size, i.e. it is proportional to the square of the number of steps along each side.

This concept of a three-dimensional scan can be immediately applied, assuming time is the third dimension. For then, as bandwidth is reduced, not only spatial but also temporal resolution is reduced which is exactly what is required if objects move at a given speed in terms of picture dimensions per second provided they are not tracked by the eye. Moreover, the loss is shared out equally amongst the three dimensions so that, compared with the two-dimensional case, the spatial resolution decreases more slowly.

There are, however, certain practical limitations in this approach. The first, and most fundamental, is that the time dimension is unbounded unlike the two spatial dimensions. Thus the scanning of an arbitrary sequence of images must consist of a sequential *series* of fractal cubes. If the starting and finishing points of the intra-cube scans are chosen to correspond to different times with the same place then the discontinuities caused by abutting the cubes may be minimised.

Secondly, because the scan moves forwards and backwards in time, the technique implies that the image sequence is stored in some way before being scanned. If the imaging device is discrete it is quite feasible to build storage into it so that successive images are stored as planes of samples, ready for scanning. But clearly there is a limit to the number of such planes. Similarly, corresponding discrete display devices with storage could be developed. In such a situation, however, the display of the images would be subject to a delay, apparently equal to twice the temporal dimension of the fractal cube. This is because the information at the beginning of the cube cannot be sent until the information at the end has occurred and the information at the beginning cannot be displayed until the information at the end has been received. This delay can, however, be halved as the scan is symmetrical in time about the midpoint.

Thirdly, the cube of information scanned by a single curve must become a cuboid. This results from a combination of setting a desired system delay and a frame rate for a particular spatial resolution. For example, suppose the frame rate is set at 64 Hz for a 512×512 spatial resolution, corresponding approximately to the present situation, and the system delay is set at ¼ second. Then the delay must be 16 frames, giving a $512 \times 512 \times 16$ cuboid and a spatio-temporal 'aspect ratio' of 32. It is this quantity which remains constant as the scan order changes. Thus, the next lowest order would have $256 \times 256 \times 8$ scans with a frame rate of 32 Hz whilst the next highest would have $1024 \times 1024 \times 32$ scans with a frame rate of 128 Hz. This all assumes that there exist one or more subsections of the three-dimensional curve which scan a complete cuboid with the assumed spatio-temporal aspect ratio. On the evidence of Fig. 6 this seems quite likely.

If the foregoing ideas seem a little too radical then, at least, the concept of three-dimensional scanning can be applied to the transmission of conventionally scanned images to optimise the use of bandwidth, using raster-to-Peano and Peano-to-raster converters at either end of the link.

8. PRACTICAL EXPERIMENTS

In order to answer some of the questions raised in Section 5, pictures were Hilbert-scanned and the resulting signal displayed using a continuous scan. The vehicle used for these experiments comprised a digital *RGB* picture store with analogue video outputs feeding an oscilloscope, capable of being accessed by a VAX computer.

8.1 The basic method

The VAX computer contains still and sequence picture material as files. The order of data in the files is that corresponding to a conventional raster scan except that the two interlaced fields of a picture are stored as a single picture in vertical coordinate order. These files can be loaded into the picture store whose output is scanned at video rate on the conventional interlace raster standards to form a video signal. Clearly, then, all that is needed to simulate a Hilbert-scanned video signal is to create new files in which the data are rearranged so that when loaded into the store and read in the conventional way the data appear as if obtained in Hilbert scanning order. In addition, to display the signal, the x and y coordinate waveforms are needed. These can be computed, stored as picture files and loaded into the store to provide deflection signals in synchronism with the video signal, taking advantage of the fact that the *RGB* store

provides three simultaneous outputs. The video signal is loaded into the green store, the x deflection signal is loaded into the blue store and the y deflection signal is loaded into the red store. The x and y deflection waveforms are taken to the corresponding inputs of the oscilloscope, operating in X-Y mode and the video signal is taken to the Z MOD input to modulate the oscilloscope trace brightness.

A limitation of this approach is that, as the digital-to-analogue converters on the output of the stores have only eight bits, the resolution of the coordinate waveforms is limited to 256×256 points. A further consequence of this simulation is that, as the store outputs are conventional video signals, all three waveforms will contain blanking intervals during which the signal will take the value of black level. This means that the x and y deflection will periodically return to a fixed position whilst the Hilbert-scanned video is at black level. This should be of no consequence provided the display is suitably adjusted.

8.2 Preliminary experiments

The first step was to derive the algorithm describing the coordinate sequences. This was done by the method indicated in Appendix 3 and a program was written to derive each order from the previous one. The eighth-order sequence was then written as two picture files. In so doing, advantage was taken of the disparity between the allowable maximum picture file size of 720×576 elements and the actual number of points, 65536, to provide a number of options. The excess capacity amounted to a factor of just greater than six so that each point could be repeated up to six times (depending on the actual file size) in a conventional picture period. This repetition could be exploited in a number of ways. At one extreme, the point frequency could be as high as the element frequency, resulting in a high bandwidth deflection signal and a picture frequency of up to six times the conventional rate. At the other extreme the point frequency could be as low as possible, resulting in a waveform approximating to 'boxcar' point transitions and a picture frequency equal to the conventional rate. In between these extremes, sections of the scan could be repeated to simulate a high local frequency but a low overall picture frequency.

The unmodulated scan produced by these waveforms was observed and the following conclusions were drawn.

1. A picture frequency of 25 Hz is too low to support the illusion of a uniform area of brightness. A frequency of 50 Hz suffices provided random eye movements are inhibited. If not, break-up is observed as a split between the bottom two quarters of the scan.

2. At the highest point frequency, the deflection system was unable to follow the idealised curve and the result appeared somewhat like knitting. This failure could be attributed to several factors. In the first place, the digital-to-analogue converters contained filters cutting at only about 80% of half the point frequency. It was subsequently discovered that the dominant factor was the inadequate bandwidth of the oscilloscope's horizontal deflection circuit which was limited to 2 MHz, 30% of the required amount.

3. At the highest point frequency, no effects due to uneven motion could be observed.

4. At half the highest point frequency the trace squared up considerably and at a quarter of the highest frequency effects due to uneven motion began to be seen as a matrix of dots.

5. The 'runs' where no transitions occur between blocks were observable on high magnification but not on low magnification. This is a consequence of the ratio between the beam spot size and the point spacing.

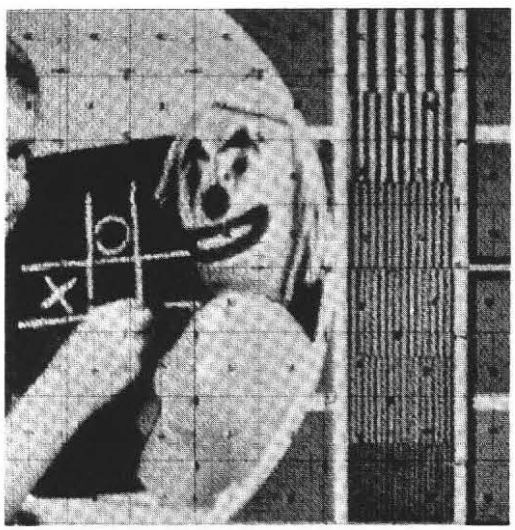
6. The whole trace appeared to 'boil' at low level in a random manner. Sometimes this cohered as blobs rotating in square cells, the direction of rotation in adjacent cells being opposite. This was later traced to breakthrough of a sinusoidal signal.

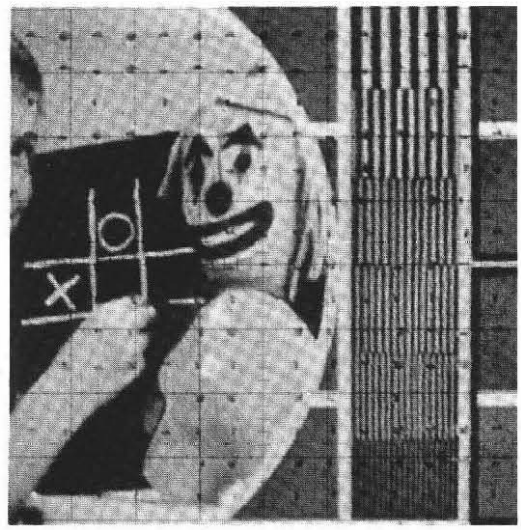
7. Correct reproduction of the whole scan required a good low frequency characteristic from the deflection circuitry. Switching the oscilloscope from DC to AC coupling caused the scan to break up into a series of rhombi.

8. The effect of blanking in the deflection waveforms caused regular patches to appear all over the trace area marking the junctions of scan sections where the scan was suspended to return to the 'black level' point.

The next step was to use the coordinate files so produced to rearrange the data in a 256×256 point picture file in sympathy with them. A window of BBC Test Card 'F' was chosen which included the frequency bars and some of the central picture material. Display of this window at a picture frequency of 50 Hz with the highest point frequency showed inadequacies in the finest frequency bars where the artifacts appeared as regular perturbations of

(a)

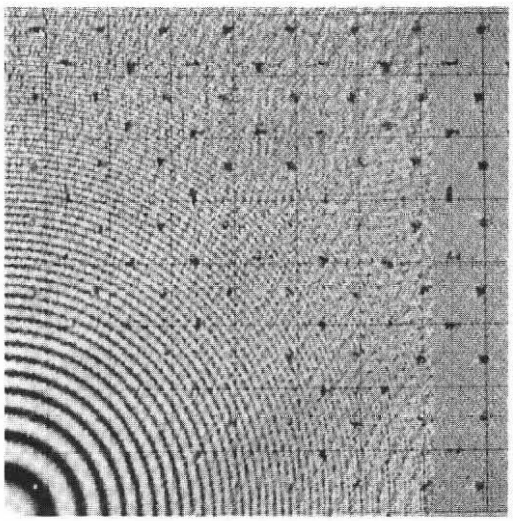


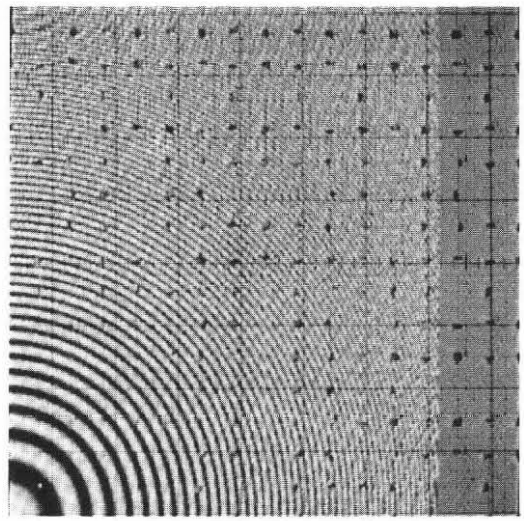


(b)

Fig. 30 - Display of Hilbert-scanned Test Card, eighth-order.
(a) Highest point frequency (b) Half the highest point frequency

(a)

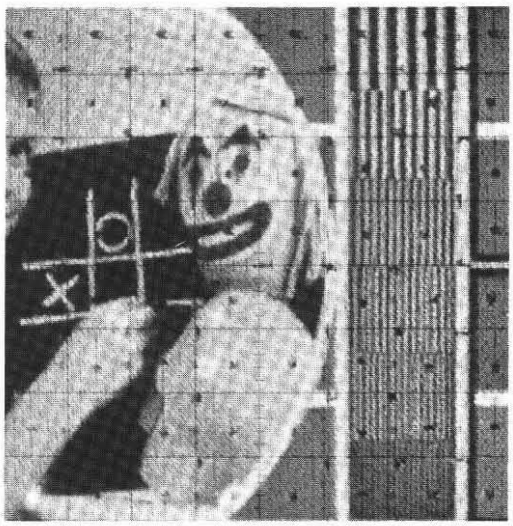


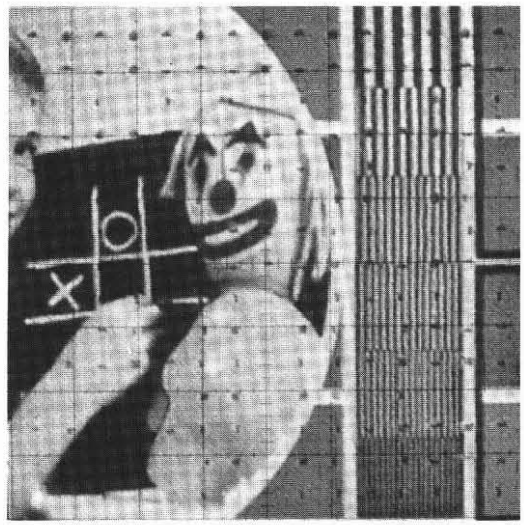


(b)

Fig. 31 - Display of Hilbert-scanned zone plate, eighth-order.
(a) Highest point frequency (b) Half the highest point frequency

(a)





(b)

Fig. 32 - Blanked interlaced display of Hilbert-scanned Test Card, eighth-order.
(a) Highest point frequency (b) Half the highest point frequency

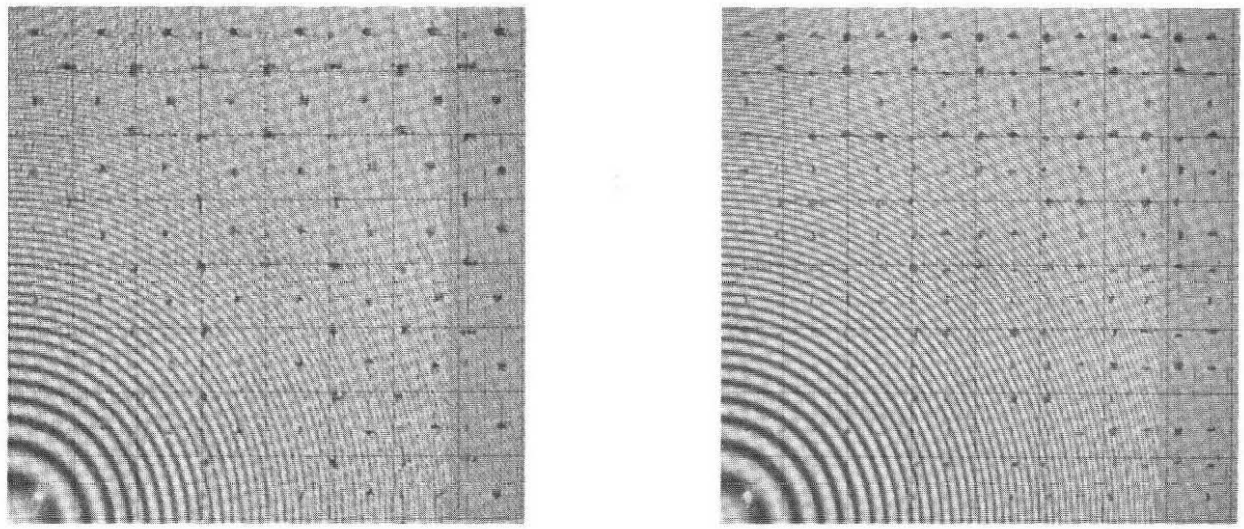


Fig. 33 - Blanked interlaced display of Hilbert-scanned zone plate, eighth-order.
(a) Highest point frequency (b) Half the highest point frequency

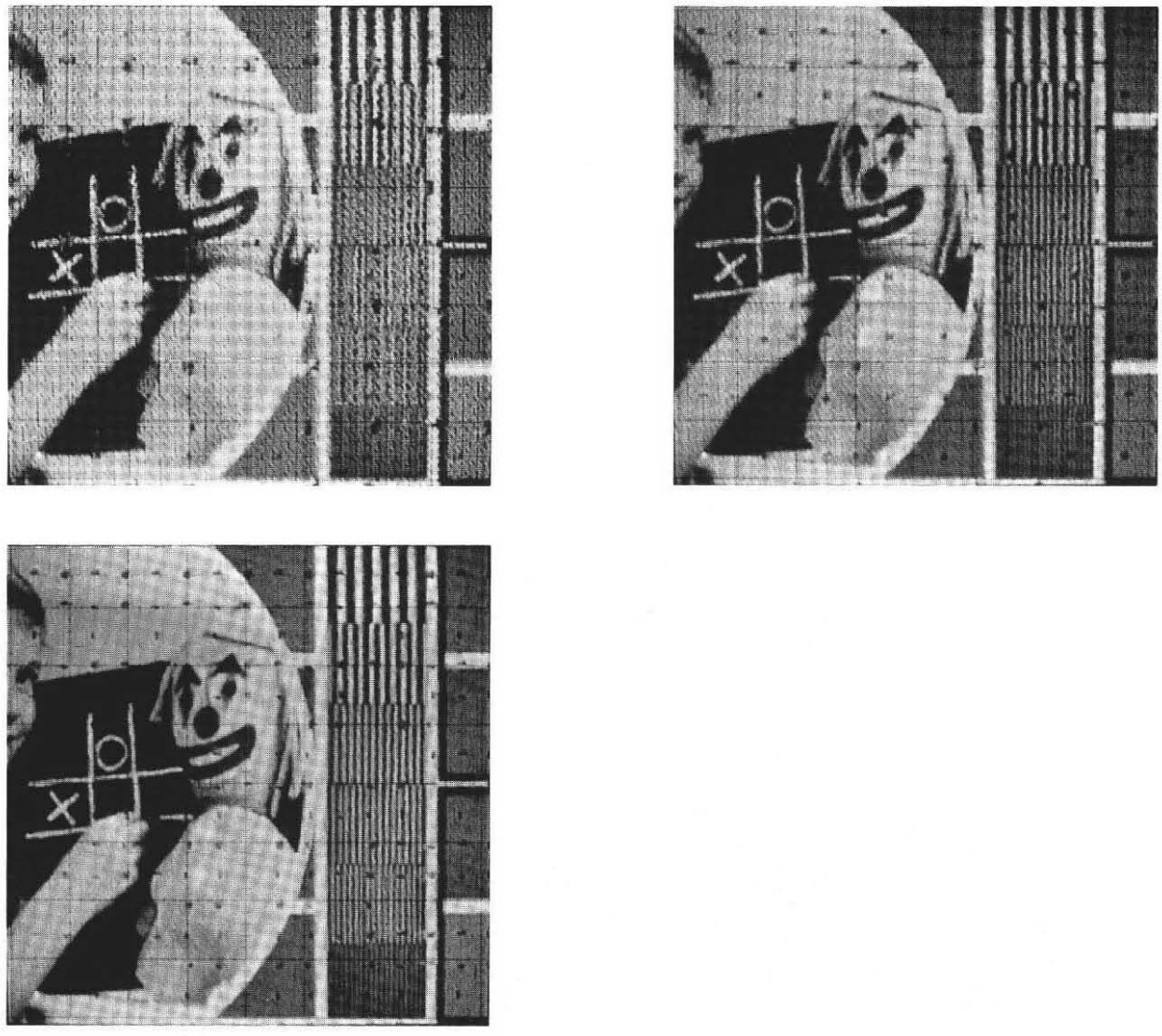


Fig. 34 - Non-blanked interlaced display of Hilbert-scanned Test Card, eighth-order.
(a) Highest point frequency (b) Half the highest point frequency (c) A quarter of the highest point frequency

the bars like weaving on a carpet. Display at half this point frequency considerably improved the quality of these bars as shown in Fig. 30 in which the effect of the 'blanking' patches can also be seen. The influence of point frequency on resolution was further tested by processing a 256 × 256 window of a zone plate, the result being shown in Fig. 31. The improvement in 'granularity' in going to the lower point frequency is clearly seen.

The next step was to introduce interlace. Choosing the blanked method allowed the blanked video data to be used with the previously derived sequential scan waveforms. Fig. 32 shows the effect on the Test Card whilst Fig. 33 shows the effect on the zone plate although, clearly, the temporal effects cannot be appreciated. These were hardly visible on the test card but appeared as a flickering pattern on the zone plate, centred on the diagonal frequency corresponding to the combination of horizontal and vertical point spatial frequencies. The insertion of the blank samples naturally reduced the contrast of the display, giving it a washed-out appearance.

The non-blanked method of interlace was also looked at for comparison. This required new scan files as well as picture files. Fig. 34 shows the effect on the Test Card. It can be seen that, at the highest point frequency, the result is poor compared with the blanked case but reducing the point frequency by a factor of four gives an acceptable result. Naturally, the removal of the blank samples restores the contrast to that in the sequential mode.

The next step was to smooth the scan waveforms to demonstrate downwards compatibility. Fig. 35 shows the display of the Test Card with the scan smoothed by various powers of two. The smoothing was done by taking a running average over a given number of points. Care had to be taken to ensure the correct phasing between the smoothed scan and the video signal by assigning the average to the

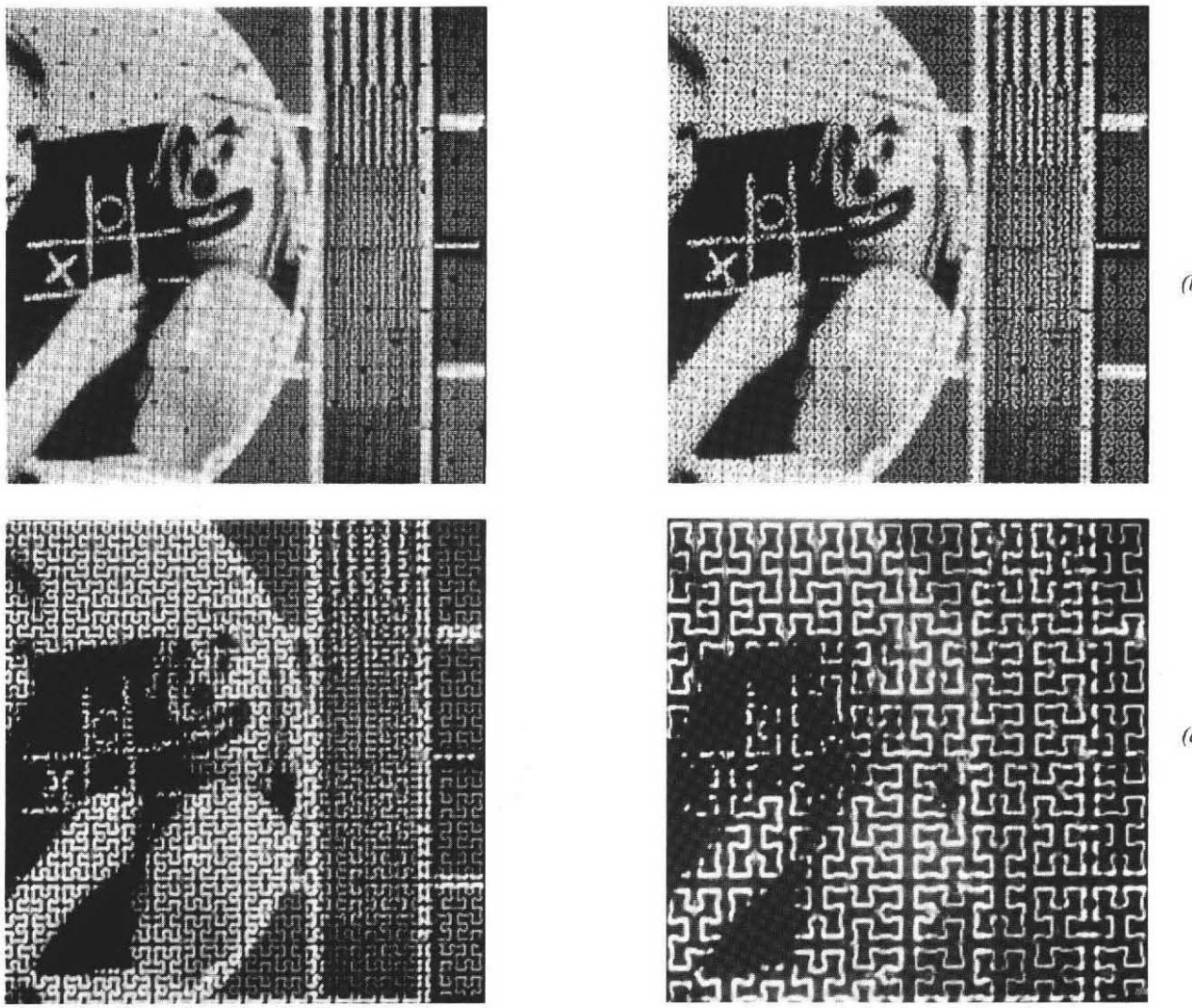


Fig. 35 - Display of Hilbert-scanned Test Card, display smoothed by various factors.
(a) 4 (b) 8 (c) 16 (d) 64

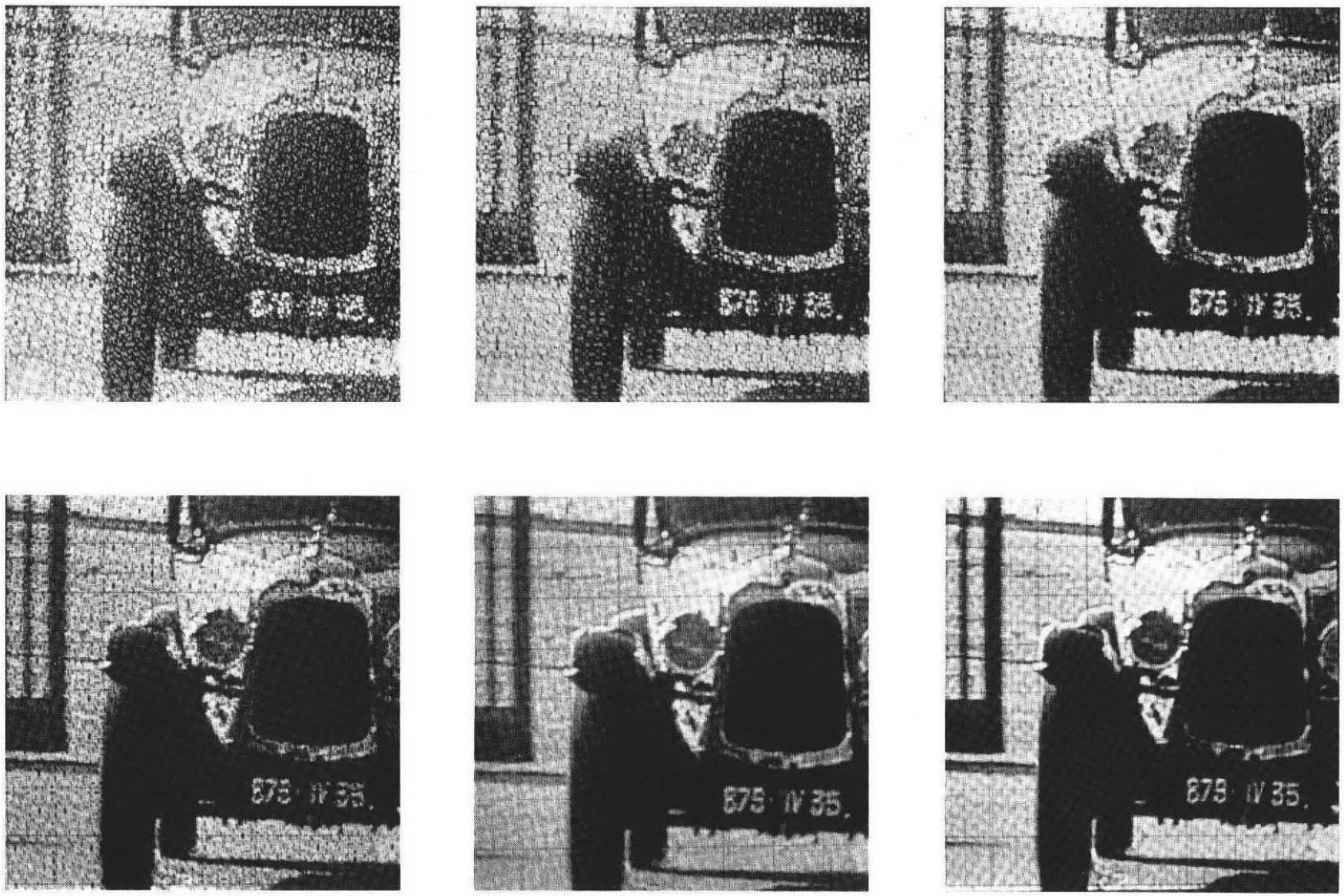


Fig. 36 - The build-up of a single frame from the displayed sequence of 'Voiture' scanned by a 12th-order interlaced 300 Hz Hilbert scan.

correct location. Note that the smoothing by an odd power of two causes the scan to change shape completely. In this series of pictures the display spot was left unchanged to emphasise the nature of the scan. It could be argued, however, that the spot should increase in size in sympathy with the point spacing to simulate a coarser display aperture. Nevertheless, the picture material is recognisable even with the coarsest display which corresponds to a 64-fold reduction in bandwidth.

The final step was to investigate higher-order interlaced scanning. Non-blanked scanning was chosen to investigate 12th-order interlaced scanning with a 300 Hz field rate, 300 being the LCM of 50 and 60. This choice also allowed each point to be stretched to six clock periods, giving a good approximation to a 'boxcar' waveform and simulating a discrete display quite well. A $256 \times 256 \times 4$ window of the 'Voiture'* sequence was chosen as the picture material. No untoward effects could be discerned as a result of the high order interlace but the increased field rate helped greatly in avoiding the 'split' effect caused by random eye movements, mentioned earlier. Fig. 36 shows the build-up of a single frame from the displayed sequence.

8.3 Hardware generation of the scan

In order to eliminate the artifacts caused by line blanking in the computer simulation, and to discover problems that might arise in a Hilbert scan of more than 256×256 points, it was thought useful to build a hardware Hilbert scan generator, capable of producing digital and analogue outputs. One way of deriving the scan would be to use a memory containing x and y values, addressed by a counter, the two outputs of the memory being available directly to address a discrete scanning device, or fed into two digital-to-analogue converters to produce continuous waveforms for a continuous scanning device, as shown in Fig. 37. The contents of the memory could be derived according to the method described in Appendix 3.

The drawback of this method is that the memory becomes unacceptably large as the order of scan increases. An nth order scan would require a memory of 2^{2n} locations, each of $2n$ bits, containing n-bit x and y values. Therefore the generator was based on a totally different algorithm derivation[8], given in Appendix 4. The design† was capable of generating a scan up to ninth order, giving 512×512 points.

The generator was designed either to accept an external clock or to free run with an internal clock frequency of 13.5 MHz. In the latter case, the resulting picture frequency was therefore approximately 51.5 Hz, being the clock frequency divided by 2^{18}. The form of the scan was a modified Hilbert curve wherein the first and last quarters were reflected about 45 degree axes so as to mirror the second and third quarters and enable the curve to be endless as shown in Fig. 38.

The ninth bit in the scan positioning could not be resolved until monolithic DACs were substituted for discrete-component DACs. Even then, to obtain a satisfactory scan at the highest resolution, great care had to be taken over physical layout to minimise clock breakthrough. It was found necessary to bypass the filters in the DACs and insert low-pass filters cutting at 11 MHz. In addition, the output was contaminated by mains hum which caused adjacent areas of the scan to overlap, varying at the beat frequency of the mains and picture frequencies. This was cured by upgrading the power supplies. Residual overlap was caused by the large DC offset of the waveforms which was a fault of the DACs. Fig. 39 shows a magnified portion of the best scan obtained. The clear disparity between horizontal and vertical performance is caused by the lower bandwidth of the horizontal deflection system.

In order to enable the scan to be modulated, a 'seamless' video signal, without blanking, had to be available, synchronised to the generator picture phase. This could not be done until a higher capacity *RGB* picture store became available, capable of being configured in a seamless way. Then it proved convenient to lock the generator to the store by adding a synchronising circuit to the generator. This used a pulse produced by the store, occurring at picture frequency, to load the counter in the generator with a preset value, thereby allowing the synchronising phase relative to the video to be adjusted. Means were also provided for finely adjusting the relative timings of the x and y deflection signals. Fig. 40 shows the picture quality that was obtained. In comparing this with Fig. 30 it must be borne in mind that the ratio of the spot size to the detail is doubled and so the spot would have a significant effect.

8.4 Scan smoothing

Scan smoothing was investigated more closely. This was done firstly using hardware low-pass filters on the output of the hardware generator, taking advantage of the absence of the blanking intervals. However, it proved more flexible to revert to computed scan files, filtered by software filters. This had the additional advantage of ensuring correct synchronisation with the video. Both simple averaging and sharp-cut filtering were tried and it appeared that there was very little difference between the two,

* See Section 10.

† The generator was designed by J.P.Chambers.

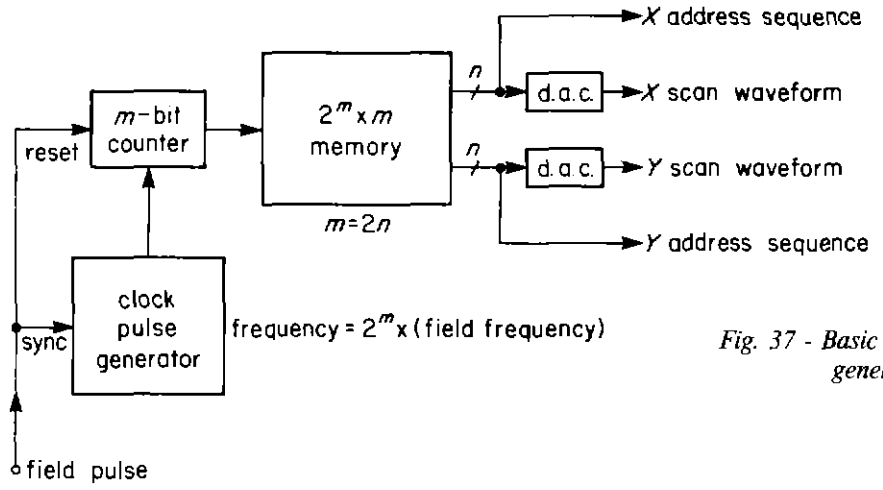


Fig. 37 - Basic circuit for one method of generating the scan.

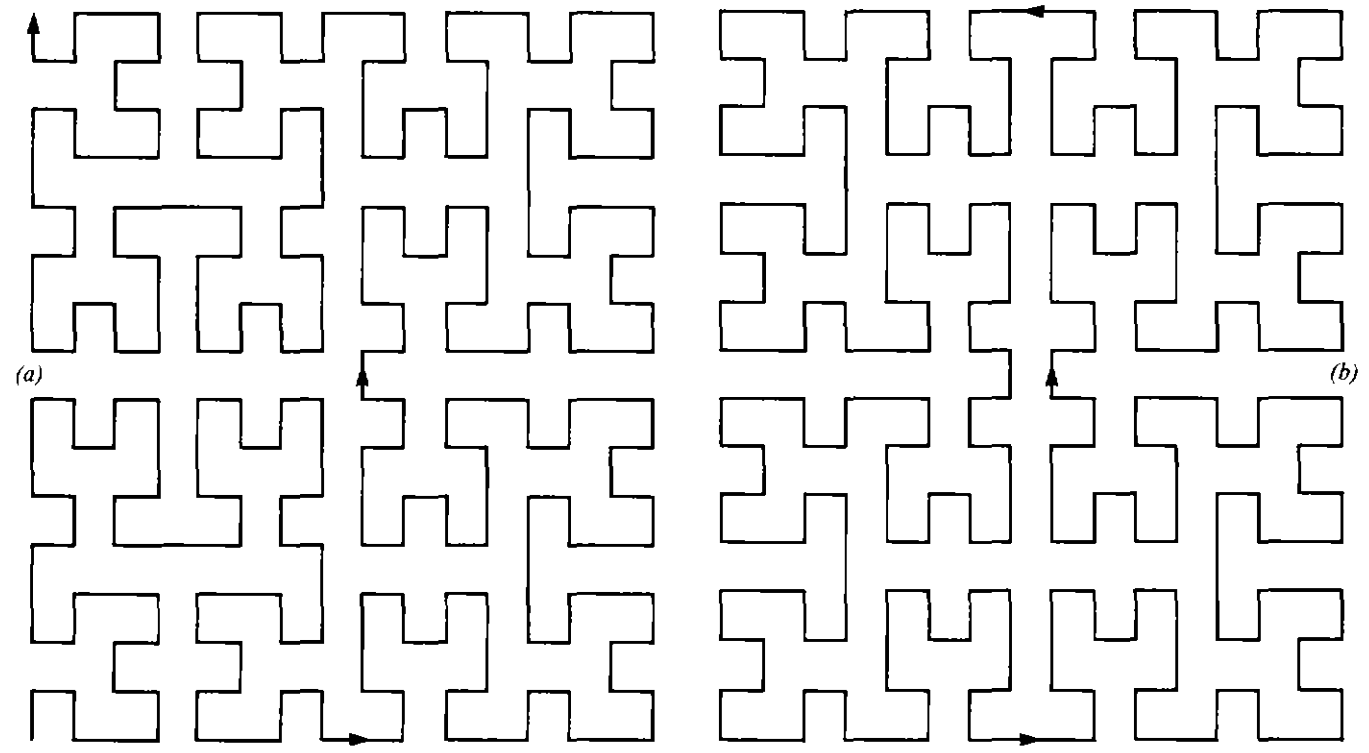


Fig. 38 - The form of the endless scan produced by the hardware scan generator.
(a) Conventional type (with field flyback) (b) Continuous type (no flyback)

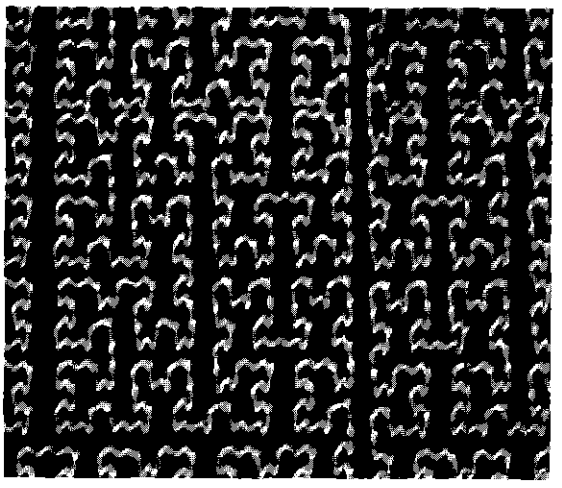


Fig. 39 - Magnified portion of the generator scan in the highest resolution mode.

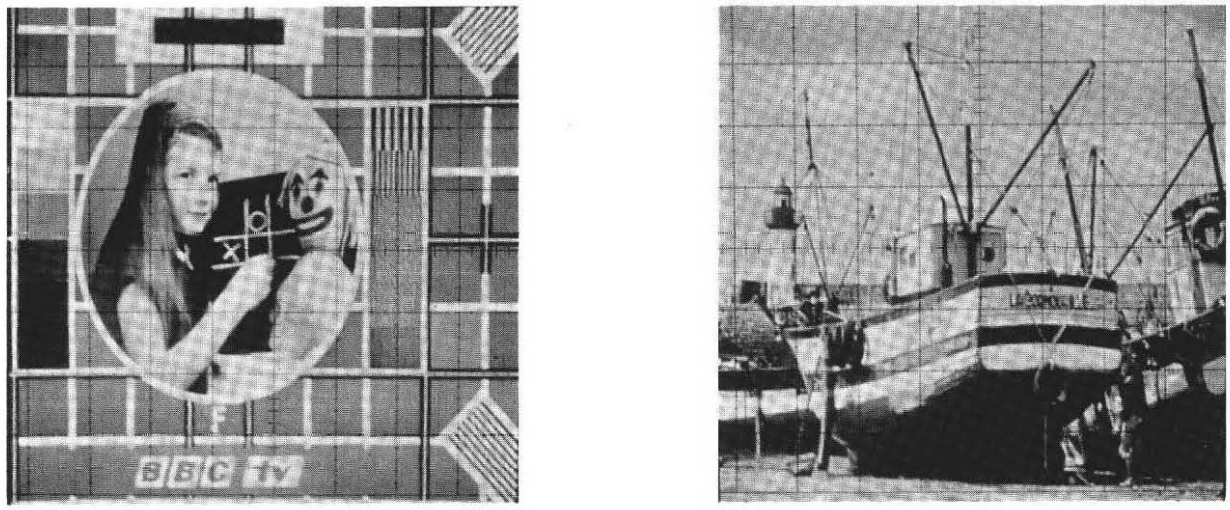


Fig. 40 - 512 × 512 pictures obtained with hardware scan generator.

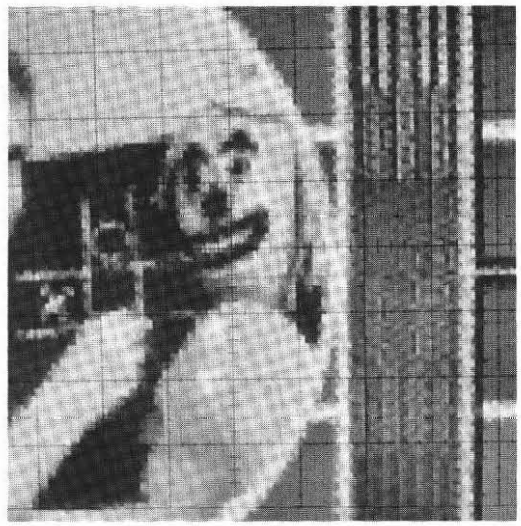


Fig. 41 - Display of Hilbert-scanned Test Card, video smoothed by 16.

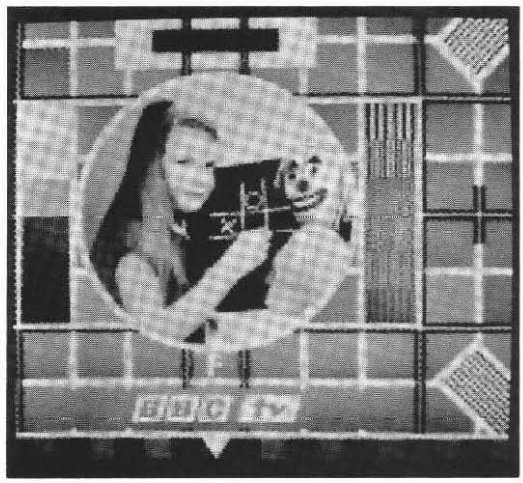
although the exact scan shape varied slightly. By defocusing the oscilloscope the spot size could be increased to hide the scan structure and give a more acceptable picture and a better representation of a lower resolution display.

8.5 Video smoothing

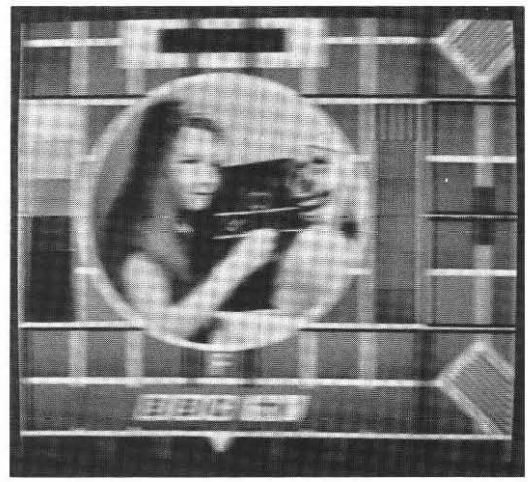
The video signal was smoothed, either by simple averaging or by sharp-cut filtering. It appeared that sharp-cut filtering gave a similar result to averaging but somewhat sharper. Fig. 41 shows the effect of averaging by a factor of 16. Notice how edges (especially diagonal ones) appear rather jagged, the whole taking on a pixellated appearance. This occurs because of the shape of the Hilbert curve. In a given small period of time the scan tends to stay within a particular square block. Smoothing the video signal will cause the picture material *within the block* to merge together but not merge with the material in adjacent blocks which may be scanned at quite different times. The more the disparity between the signal bandwidth and the display, the more noticeable the effect becomes. By defocusing the picture one can hide the effect but it is arguable whether this blurred picture looks better.

To simulate smoothing in a Hilbert-scanned channel with raster display, the video signal was converted to raster form and displayed on a conventional monitor. This enabled comparison with conventionally transmitted pictures. Comparison with a conventional picture having the same video bandwidth (i.e. smoothed only horizontally) showed the Hilbert picture to be considerably better to look at, thus confirming that Hilbert scanning offers a usable method of splitting the available bandwidth equally between the vertical and horizontal directions. However, the Hilbert picture did look inferior to a raster picture which had been smoothed isotropically to contain the same total information. Fig. 42 shows these comparisons.

(a)



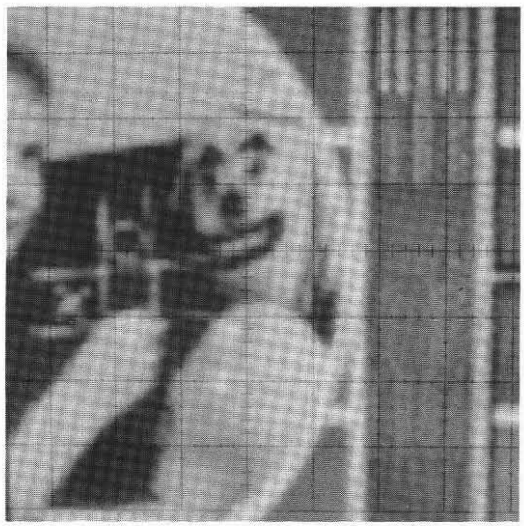
(b)



(c)

Fig. 42 - Comparison of bandwidth-reduced Hilbert scan with raster scan.
(a) Hilbert reduced by 16 (b) Raster reduced horizontally by 16 (c) Raster reduced horizontally by 4 and vertically by 4

(a)



(b)

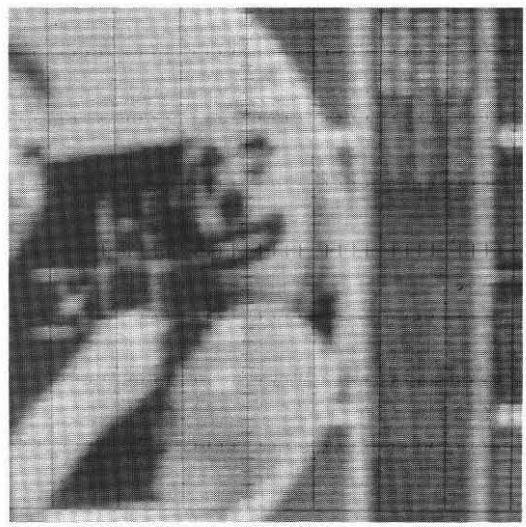
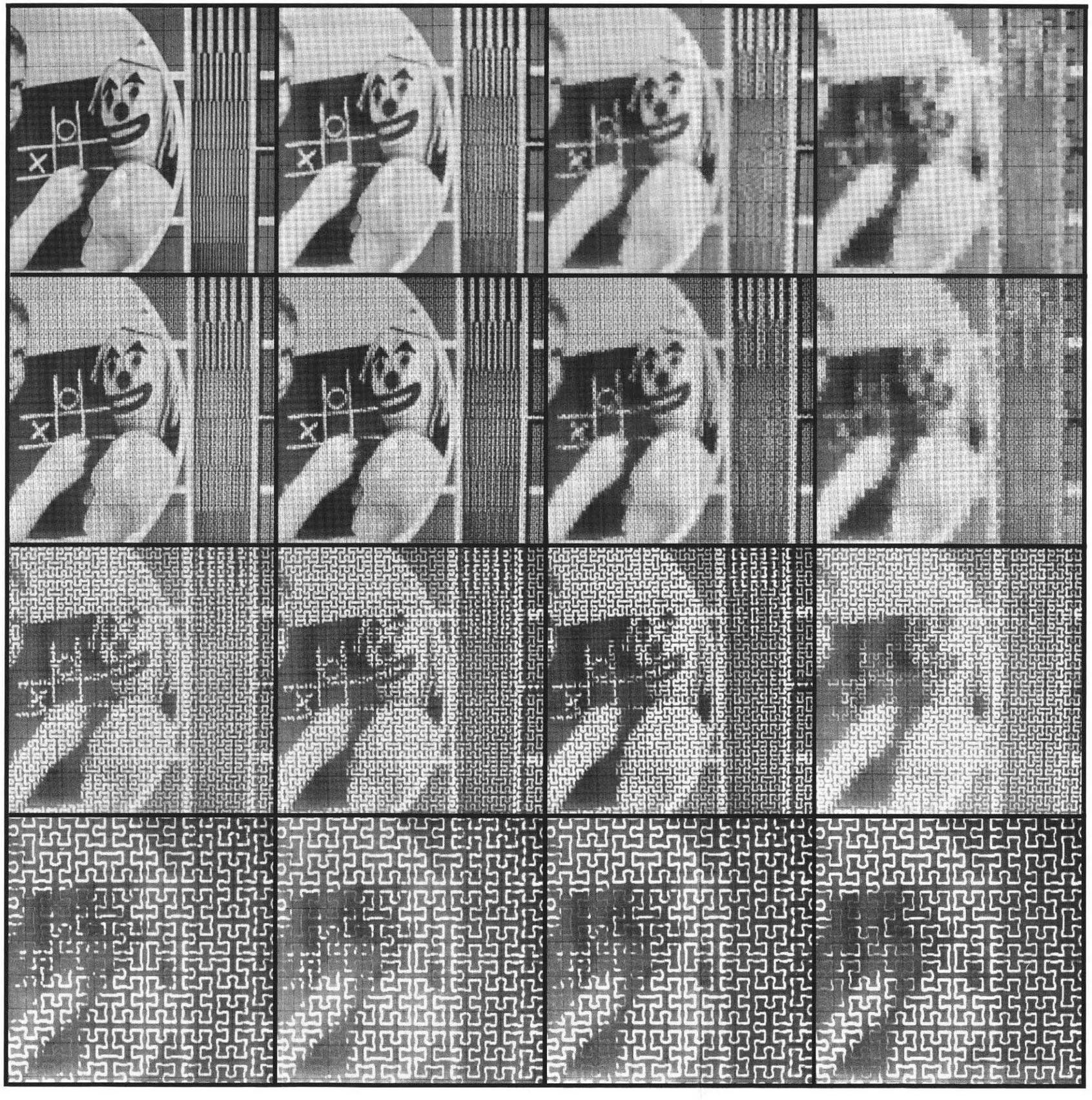


Fig. 44 - Comparison of display averaged and signal averaged situations with display defocused.
(a) Signal averaged (b) Scan averaged



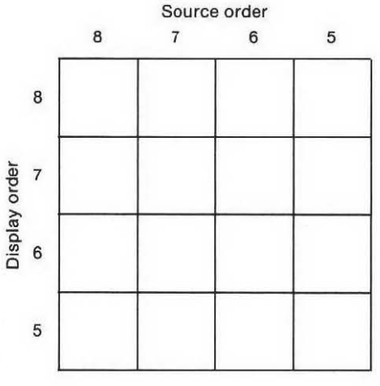


Fig. 43 - Matrix of various source and display resolutions showing upwards and downwards compatibility for sequential scanning.

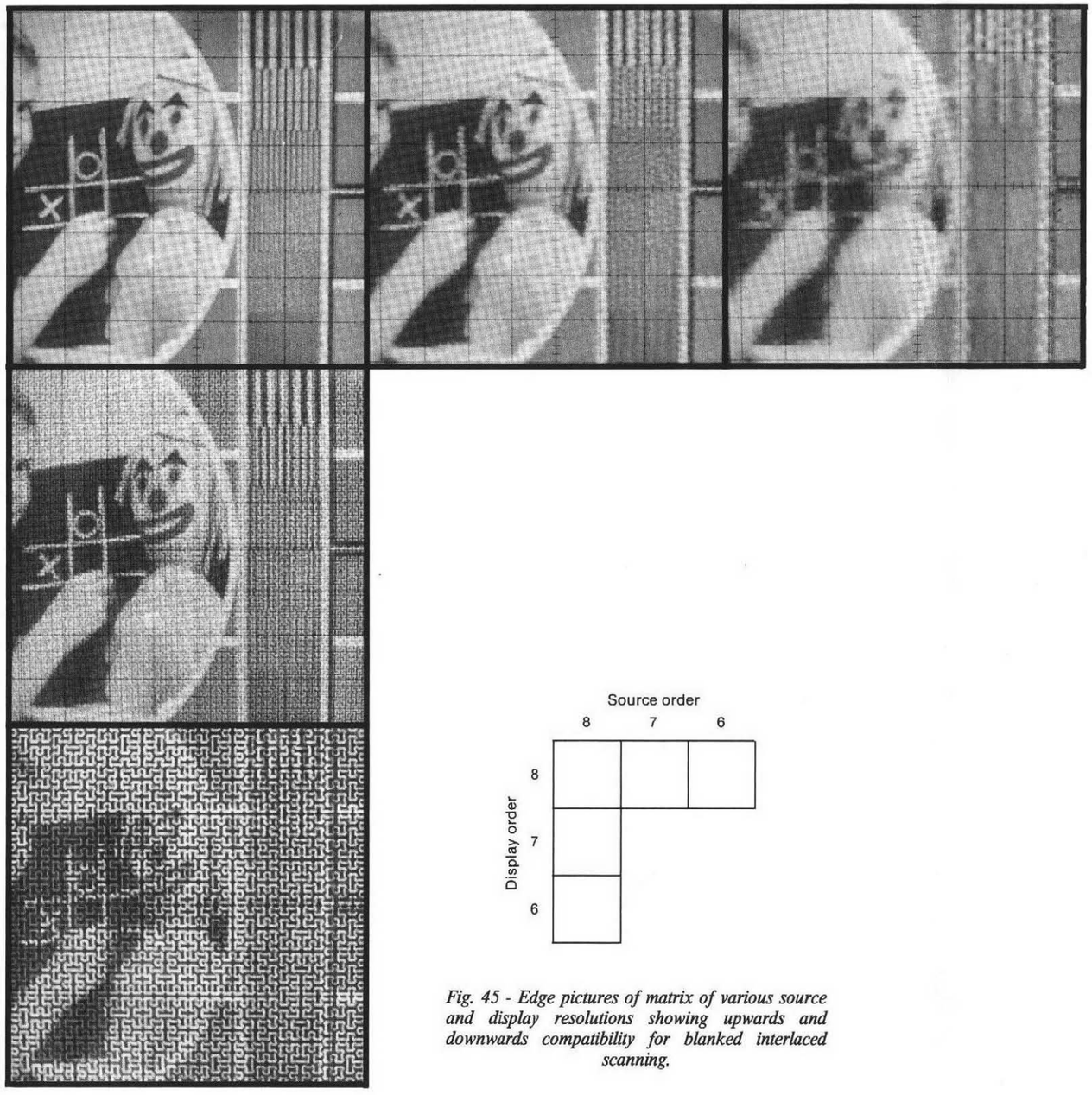


Fig. 45 - Edge pictures of matrix of various source and display resolutions showing upwards and downwards compatibility for blanked interlaced scanning.

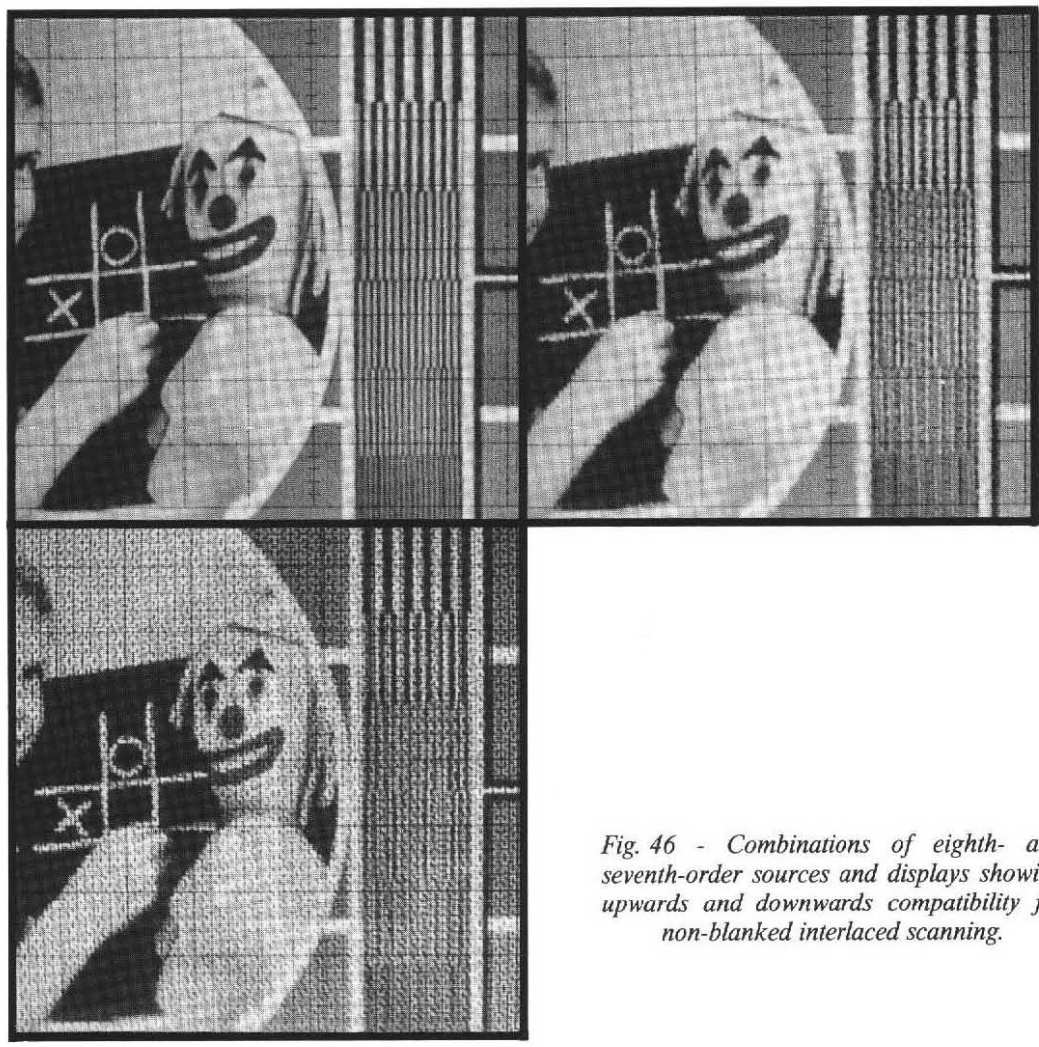


Fig. 46 - Combinations of eighth- and seventh-order sources and displays showing upwards and downwards compatibility for non-blanked interlaced scanning.

8.6 Upwards and downwards compatibility

To demonstrate upwards and downwards compatibility, seven picture files and seven pairs of scan files with bandwidth reduced (by averaging) by factors of 1, 2, 4, 8, 16, 32 and 64 were produced. By loading each of these 14 files into the *RGB* store seven times and reconfiguring to *RGB*, 49 pictures were obtained, being the various combinations of each video signal with each scan. The pictures corresponding to smoothing by even powers of two form a 4×4 matrix, Fig. 43, which shows how pictures from sources with four different resolutions appear on four different displays.

According to the hypothesis, the picture quality should be dependent on the coarser of the source and display. Thus, the diagonal in Fig. 43 forms a boundary to the left of which the quality should remain the same along a horizontal line or above which the quality should remain the same along a vertical line. In practice, the subjective quality does vary slightly above the diagonal but it is difficult to balance the impairments caused by pixellation against the improvement caused by reduced structure visibility. Fig. 44 shows a comparison between the video averaged by 16 and the scan averaged by 16 when the display spot is defocused to hide the pixellation in one case and the scan structure in the other case. Now there is little to choose between them.

Both blanked and non-blanked second-order interlaced scanning were further investigated, taking advantage of the increased picture storage to rearrange a complete picture into two adjacent stores. Combinations of averaged source and scan waveforms were made as for sequential scanning. Fig. 45 shows a compatibility matrix for blanked scanning. This is similar to Fig. 43 except that the contrast is lower and signal averaging has a more marked effect. Fig. 46 shows the compatibility between seventh and eighth orders for non-blanked interlace. As expected, the contrast is restored to that of the non-interlaced scan. Comparison with non-interlaced pictures reveals that, for example, reducing the video bandwidth of an interlaced picture by a factor of four has a similar effect to reducing the video bandwidth of a non-interlaced picture by a factor of eight. In other words,

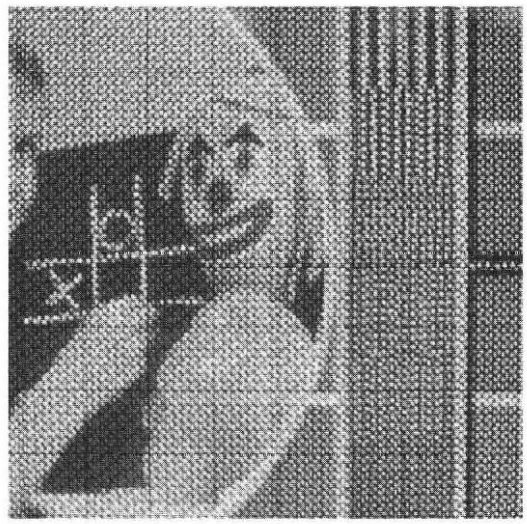


Fig. 47 - Display of a single field of the interlaced Hilbert scan.

using second-order interlace and halving the video bandwidth have a similar effect on picture definition, once the definition of the source is less than the definition of the display.

Looking at just one field of the two, fine diagonal lines are noticeable on pictures as shown in Fig. 47. This is because on either field, the scanned points form a chequer-board pattern. Fig. 48 shows a magnified portion of one field of the non-blanked interlaced scan. When both fields are displayed the picture looks rather 'lively'.

8.7 Moving sequences

Moving sequences were investigated with both hardware- and software-generated scans. During this investigation a new effect came to light which was

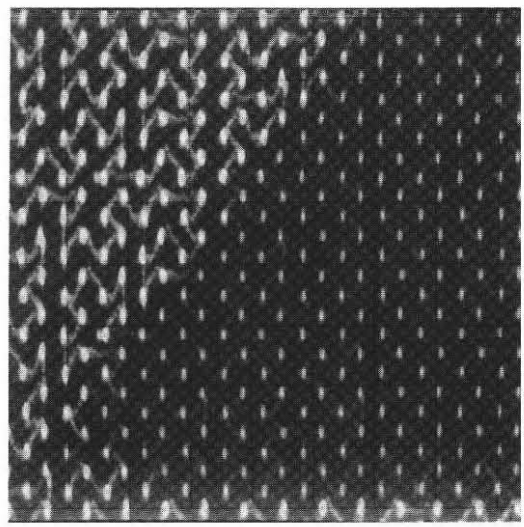


Fig. 48 - Magnified portion of a single field of a non-blanked interlaced Hilbert scan.

first noticed with the moving gate in the 'Voiture' sequence using an endless scan. When a vertical edge which crosses the horizontal line X-X, shown in Fig. 49(a), moves horizontally across the picture it appears to have a break in it, as shown. This occurs because the display scans the two areas adjacent to the line X-X close together in time but the picture information on either side of the line belongs to different frames of the original raster-scanned sequence. Because the new picture information is arranged to start at the centre of the scan Fig. 38 shows that the effect will occur at X-X. By starting the new picture in the corner, breaks occur, instead, along the line Y-Y which marks the new boundary between old and new pictures, as shown in Fig. 49(b). Starting the picture information at other points just moves the problem elsewhere. By using the true Hilbert scan the problem is reduced because there is not such a distinct boundary but the effect can still be noticed where neighbouring blocks of the scan structure meet. Compared with the endless scan, the true Hilbert scan seems to show less flicker for a given scan frequency.

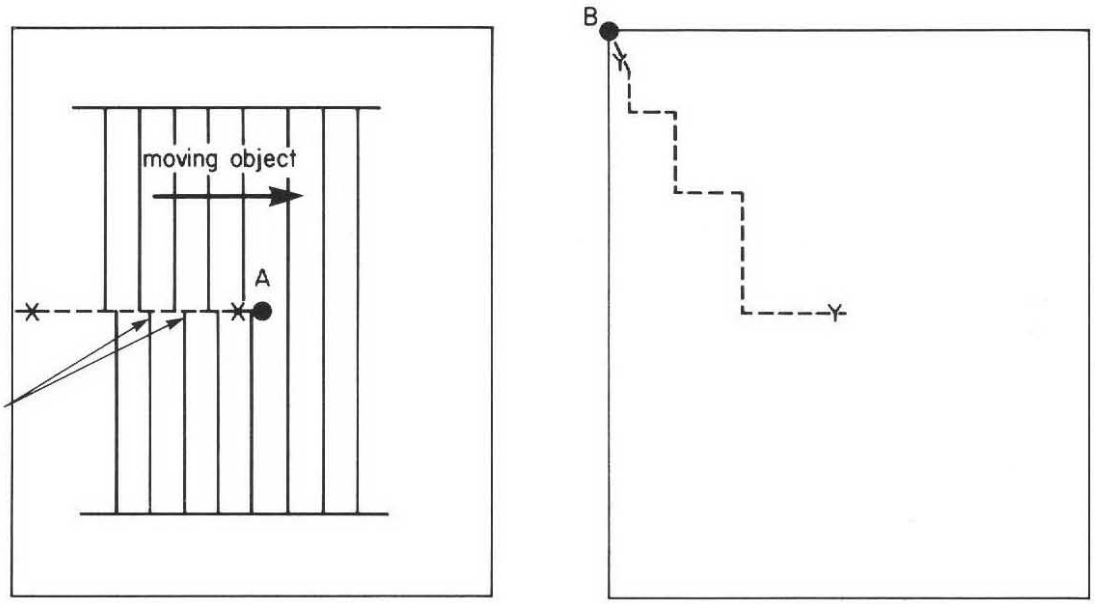
This effect is only a problem if the source and display use different scanning methods. If both the source and display are Hilbert scanned then it will not occur. Nor does it occur using Hilbert scanning only in the channel with raster source and display. With discrete devices the problem can be eliminated by ensuring that all points in a given frame correspond to the same instant of time. This is normally the case with discrete devices anyway. Then the order of sending the points has no effect on the motion portrayal.

8.8 Timing

A drawback of Hilbert scanning that became apparent is that it is far more susceptible to small timing discrepancies between video and scan than is a raster scanned picture. With the latter, delay causes a simple shift to the right whereas with Hilbert scanning, delay causes localised distortion of picture detail, the larger the delay, the less local this becomes. A one pixel delay is quite easy to see and a delay of one quarter of a pixel is just noticeable. It is thus very important that the Hilbert display scan (or the scan of a converter) is accurately synchronised to the video. Fig. 50 shows the effect of misregistration between video and scan.

8.9 Noise performance

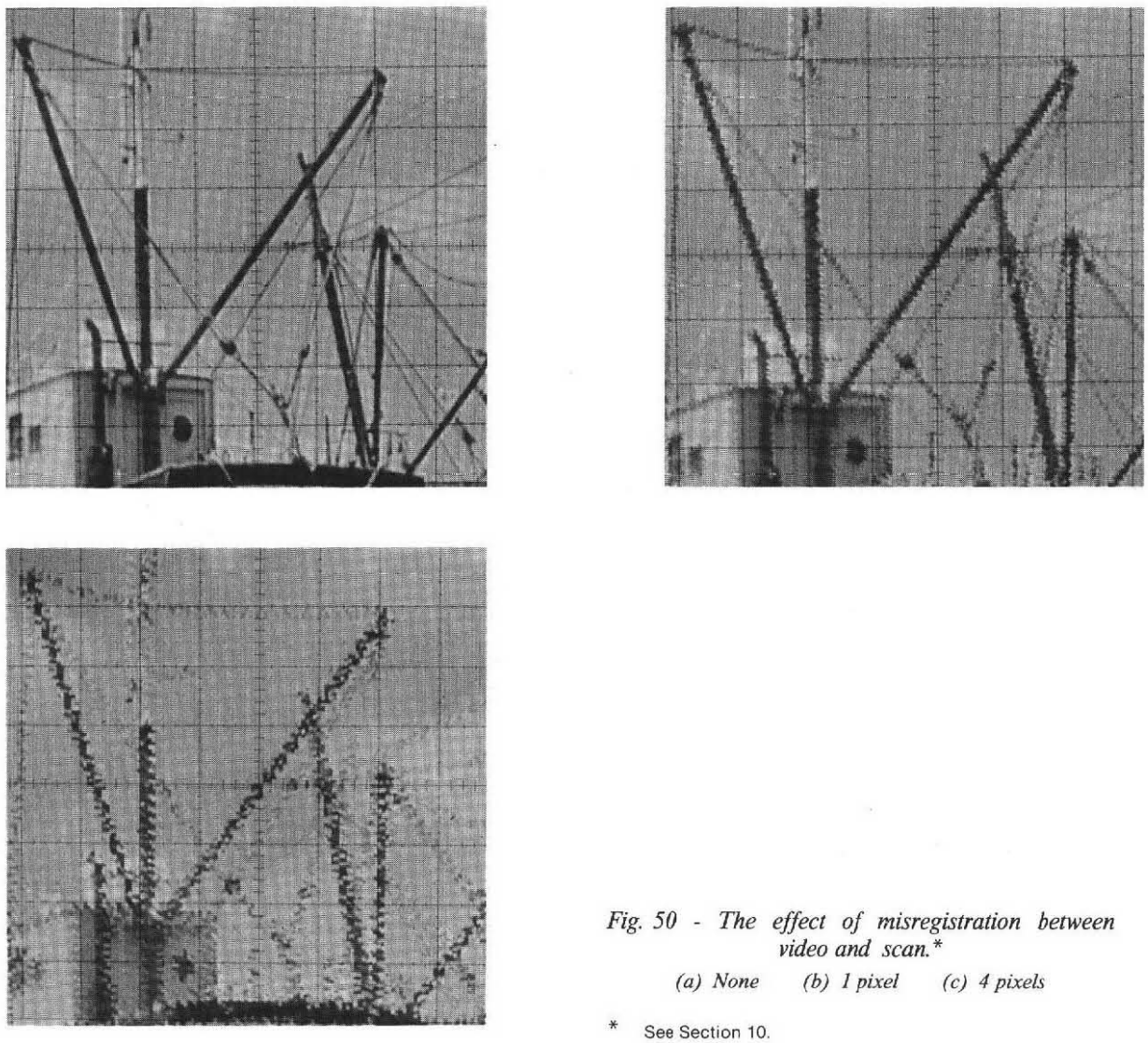
Given that the spectrum of a Hilbert scanned signal is 'pinker' than that of a conventionally scanned signal, one might expect a difference in signal to noise performance for, say, white noise introduced on the channel. This can be investigated by adding noise to the Hilbert scanned video signal and examining the



(a) (b)

Fig. 49 - The boundary between old and new picture information for the endless scan.

(a) Phased at the centre (b) Phased at the corner



(a) (b) (c)

*Fig. 50 - The effect of misregistration between video and scan.**

(a) None (b) 1 pixel (c) 4 pixels

* See Section 10.

display in either Hilbert form or converted to raster form. In practice, it proved convenient to do the latter.

The hypothesis is that when a noise signal is displayed with a raster scan the horizontal spectrum of the displayed noise is the same as the spectrum of the signal, whatever it may be, but the vertical spectrum is white (provided that the noise bandwidth is significantly greater than the line frequency). However, when a noise signal is displayed with a Hilbert scan (or converted from Hilbert to raster scan) the signal spectrum is spread out equally between the horizontal and vertical spectra of the picture.

In practice, white noise looked very similar with both a Hilbert and a raster picture. This is to be expected from the above because the spectrum of the noise signal is flat. Thus both the horizontal and vertical spectra of the displayed noise are flat for either raster or Hilbert scanning.

Triangular noise looks slightly worse with a Hilbert picture, especially when a sequence is displayed. The noise in a single Hilbert picture shows more twitter than its raster equivalent. Fig. 51 shows the computed horizontal and vertical spectra of the displayed raster noise and Fig. 52 shows the same for the Hilbert noise after conversion to raster form. The horizontal spectrum of the raster noise is similar to the signal spectrum (i.e. it is triangular) whereas the vertical spectrum is white, as the hypothesis suggests. However, with the Hilbert noise, note how the signal spectrum has been spread out in the unscrambling process equally between the horizontal and vertical spectra, again as suggested above. The horizontal noise of the raster picture is concentrated at high frequencies with none at zero frequency whereas the Hilbert picture has more noise at low horizontal frequencies in exchange for less at high frequencies. As the eye is less sensitive to the higher frequencies it finds the Hilbert picture more distracting. In the vertical direction, on the other hand, the situation is

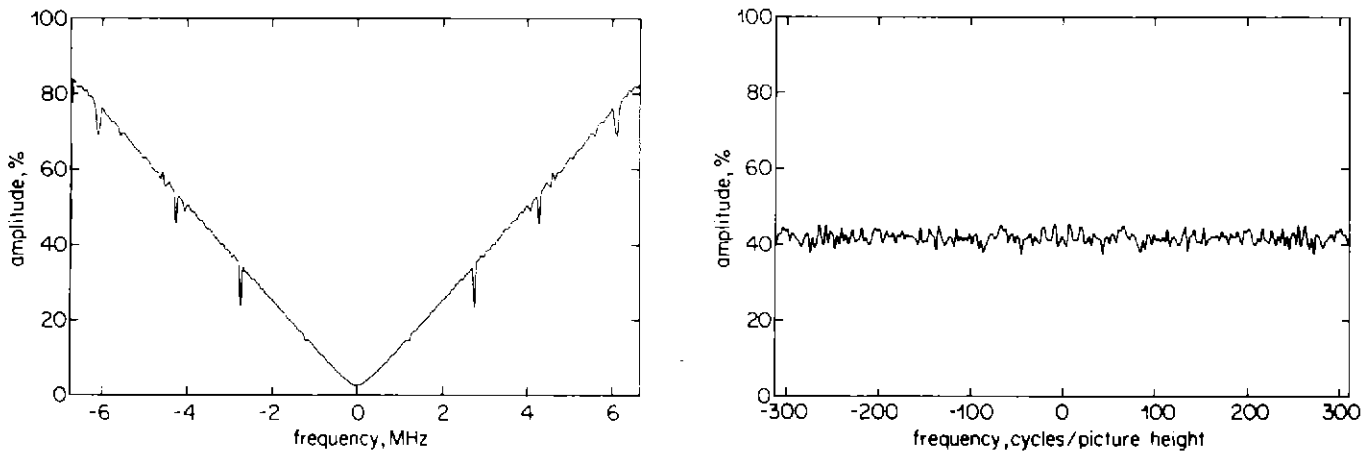


Fig. 51 - The horizontal and vertical spectra of triangular noise displayed on a raster scan.
(a) Horizontal: 256 × 256 pixel patch of FM noise. (b) Vertical: 256 × 256 pixel patch of FM noise.

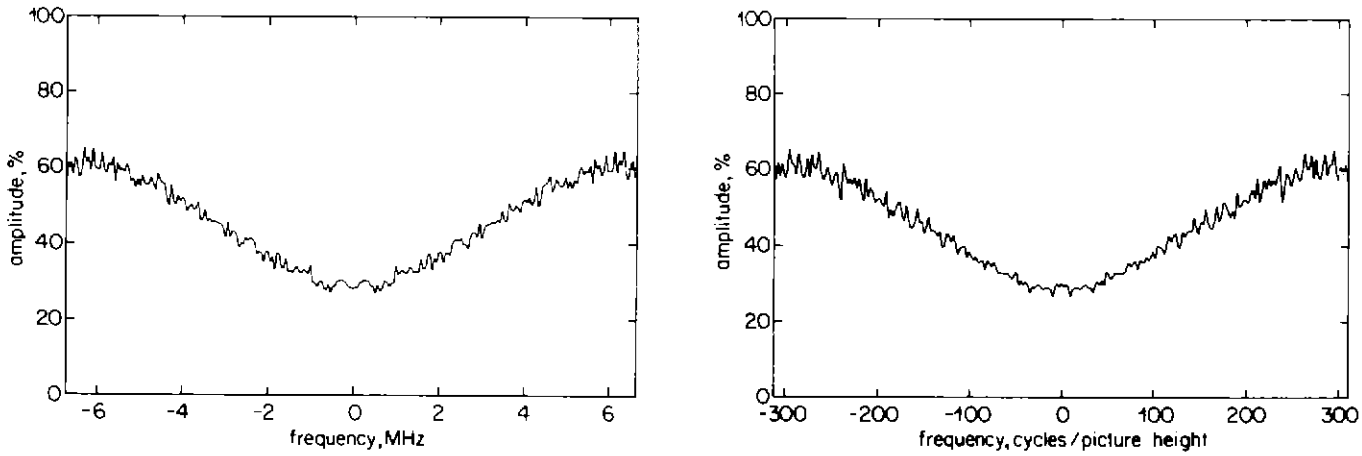


Fig. 52 - The horizontal and vertical spectra of triangular noise – Peano signal unscrambled for display on a raster scan.
(a) Horizontal: 256 × 256 pixel patch of FM noise. (b) Vertical: 256 × 256 pixel patch of FM noise.

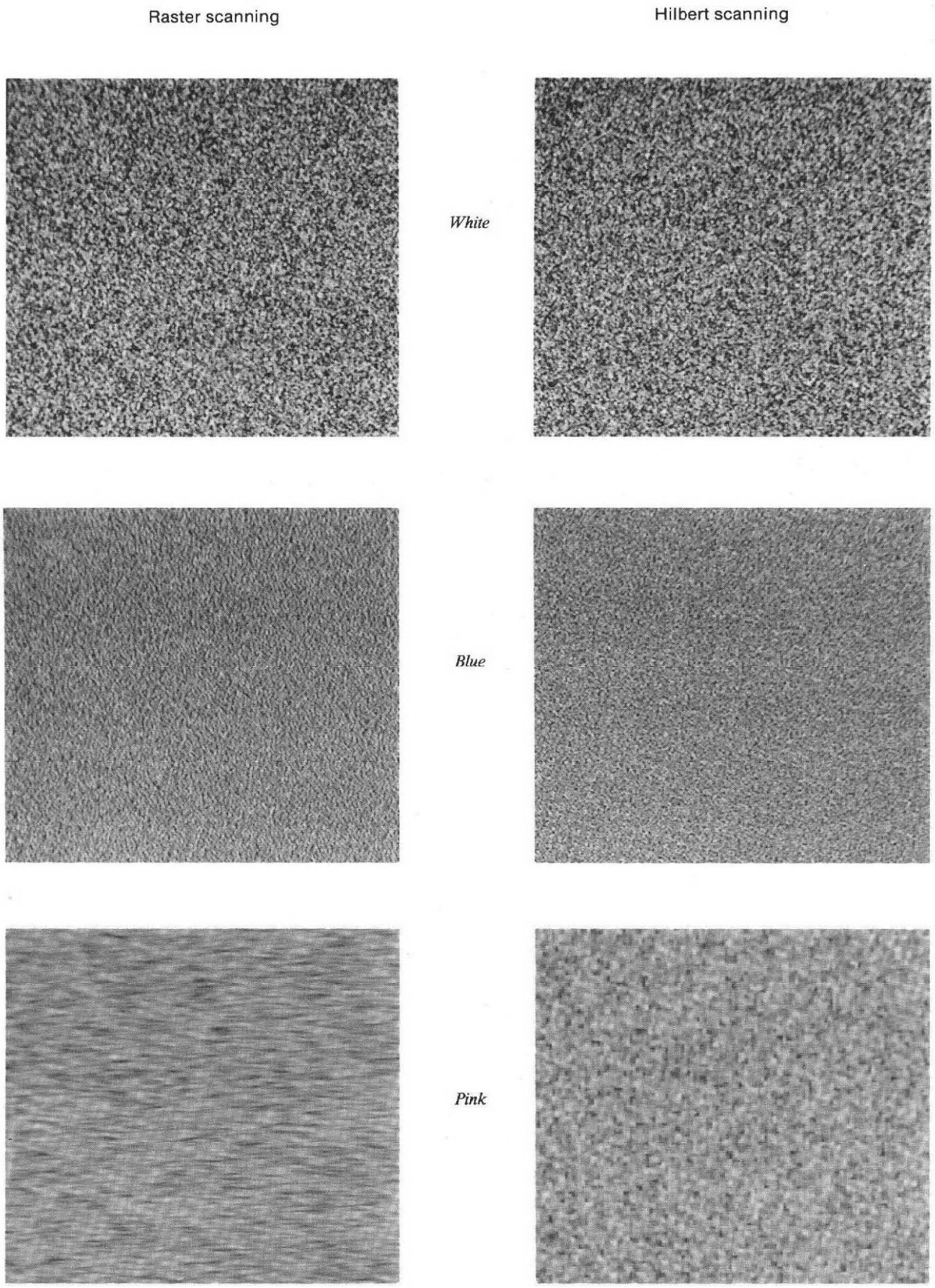


Fig. 53 - Comparison of noise in raster and Hilbert scans.

reversed with the Hilbert picture having more noise at the higher frequencies and less at the lower frequencies but now the high frequency causes twitter which makes it more visible than the low frequency noise and so, again, it looks noisier than the raster equivalent.

Pink noise looks very different on a Hilbert scanned picture. It looks as though it is made up of tiles of different shades, varying a little in size and shape and not always having clear boundaries between them. But in general they are squarish, the edges of the squares corresponding to local boundaries in the shape of the Hilbert scan. The minimum size of the squares is determined by the approximate bandwidth of the noise (it is rather like the effect of low pass filtering the video). Overall, the noise is more noticeable than with raster scanning.

Fig. 53 shows a comparison between the scans for the three types of noise.

9. CONCLUSIONS

Bearing in mind that all the work was done with continuously scanning display devices the following conclusions can be drawn.

1. It is possible to deflect a beam electrostatically in a CRT with a Hilbert scan to produce the illusion of a uniform patch if the field rate is 50 Hz or greater. However, eye movements will detect splits, principally between the first and fourth quadrants.

2. If the deflection system has inadequate bandwidth the scanning gracefully degrades to a structure which, when magnified, looks like 'knitting'. On pictures this manifests itself as granularity.

3. Given that the beam does not travel at constant speed, the brightness modulation this could be expected to produce does not seem to be a problem.

4. The deflection system needs a good low-frequency characteristic to avoid scan breakup and good DC stability and freedom from hum to resolve 512×512 pixels.

5. When pictures are displayed, horizontal and vertical edges develop a sub-pixel texture like weaving.

6. Hilbert scanning is upwards and downwards compatible in practice as well as in theory. The matrix of compatibility shows that pictures from a source are reproduced correctly on a display of different resolution. However, compatibility is not perfect with a higher definition display because the picture appears pixellated on the scale of the source, rather than smoothed. The effect is rather like zero-order interpolation but with an alternating displacement.

7. Where the source and display bandwidths are fixed but the channel bandwidth is variable, Hilbert scanning offers a compatible means of conveying pictures which uses the bandwidth optimally to ensure equal horizontal and vertical resolution. On the other hand, a signal obtained from a raster scan which is matched to a given channel bandwidth, giving equal horizontal and vertical resolutions, gives a better up-converted picture than the corresponding Hilbert scan. This is because it can be interpolated more accurately than the approximate zero-order mechanism offered by Hilbert scanning.

8. It is possible to introduce interlace to Hilbert scanning by sample selection along the curve. With continuous scanning devices this may be done either by blanking a non-interlaced scan or by discontinuous scan motion. A 12th-order interlace with a field frequency of six times the normal has been shown to work. Interlaced Hilbert scanning is also upwards and downwards compatible in practice though the pixellation effect is worse. Considered as a bandwidth reduction technique, simple second order interlace is roughly equivalent to reducing the bandwidth of a sequential scan by an extra factor of two.

9. When the source and display use different scanning methods, such as might arise if Hilbert scanning were used in conjunction with conventional scanning devices, discontinuities can appear on moving pictures. (The problem seems to be worse with an endless form of Hilbert scan.) These are caused by differing relative timings between points in the picture because, with both types of scanning, time progresses along the scan locus. The problem can be avoided if both raster and Hilbert scanning devices contain storage that enables all points of a field to correspond to the same instant of time, as in cine film.

10. Synchronisation between scan and video is critical for a Hilbert scan in contrast to the situation with a raster scan. A one-pixel

discrepancy is unacceptable. This implies that extraction of synchronising information must be very accurate.

11. The noise performance of Hilbert scanning is inferior to raster scanning. White noise looks similar but, surprisingly, both blue and pink noise look slightly worse. If FM transmission were to be used in a Hilbert system this would be important.

Most of the drawbacks of Hilbert scanning are concerned with continuously scanning devices. Further work is needed with calligraphic displays to investigate these in more detail but the results obtained so far indicate that continuous scanning is feasible.

With discrete devices the chief problem would appear to be synchronisation and further work is needed here to investigate ways of introducing synchronising information into the signal without destroying the hierarchical compatibility. The 'discrete temporal sampling' of discrete devices is particularly suitable for Hilbert or three-dimensional Peano scanning and when the technology of these devices, particularly displays, is sufficiently advanced, this is an idea that could well come of age.

10. ACKNOWLEDGEMENTS

Thanks are due to Peter Brightwell who was responsible for most of the practical work described in Section 8 following the preliminary investigations, and for valuable help in interpreting it.

The 'Voiture' and 'Boat' source data were kindly provided by CCETT, Rennes, France.

11. REFERENCES

1. UK Patent No. GB 2 137 844A. 1987.

2. STEVENS, R.J., LEHAR, A.F. and PRESTON, F.H. 1982. Data ordering and compression of multi-spectral images using the Peano scan. International Conference on Electronic Image Processing, 26-28 July 1982. IEE Conference Publication No. 214, pp 209-213.

3. UK Patent No. GB 2 139 411A. 1990.

4. MANDELBROT, B. 1977. Fractals — Form, Chance and Dimension. W.H. Freeman. ISBN 0-7167-0473-0.

5. PEANO, G. 1973. Selected Works of Guiseppe Peano. Translated and edited by Hubert C. Kennedy. Allen and Unwin. ISBN 0-04-164002-0.

6. WIRTH, N. 1976. Algorithms + data structures = programs. Prentice-Hall. pp 134-137. ISBN 0-13-022418-9.

7. UK Patent Application GB 2 215 935A. 1989.

8. UK Patent Application GB 2 215 167A. 1989.

APPENDIX 1

The Mechanism of Hierarchical Compatibility

A1.1 Discrete versus continuous scanning

In discrete scanning, the image is sensed or displayed at discrete points and the Hilbert curve then defines the sequence in which points are addressed. More exactly the points are, in reality, areas or apertures. The effect of these is, at the source, to integrate spatially, and at the display, to spread the light over an area. These are second-order effects which can be compensated.

In continuous scanning, the image is sensed or displayed by a continuously moving beam the locus of whose motion is the Hilbert curve. A beam can, however, approximate to a discrete device if the dwell time at the points is much longer than the transition time between points.

The signal from a discrete source is usually in continuous form, because the sequence of samples corresponding to each point will be filtered by the associated circuitry to a bandwidth of about half the sampling frequency. Thus the signal would not normally comprise 'held' values which would imply discontinuities and hence infinite bandwidth. The original sample values can be recovered approximately, if necessary, by resampling in the appropriate phase.

Likewise a discrete display accepts a continuous signal and effectively resamples it by displaying successive time windows of the signal at successive points. These windows act as a shutter or prefilter before the sampling and can be any width up to the pixel period. The effect is to cause the light output to be proportional to the integral of the signal over the window. This is a further aperture loss which can be compensated.

If a discrete source is coupled to a matched discrete display it might be thought that the ideal situation is where the samples corresponding to each source element are reproduced exactly by the display. This is arguable, bearing in mind the spatial integration that takes place over each element, but it is a simple yardstick by which to measure compatibility. In practice, because the two are linked by a continuous signal, a loss occurs which can be shown as follows.

Fig. A1.1 shows a signal obtained from discrete source samples *A,B,C,D,* etc. For the purposes of explanation the samples have been linearly interpolated although in practice the interpolation would be of somewhat higher order. As noted, the display integrates the signal before resampling, forming apertures centred on the times of *A,B,C,D,* etc. Suppose that the display shutter is 360 degrees so that the apertures of adjacent display pixels abut in time, giving an integration time of one pixel period. The integral is the sum of trapezoidal areas formed by the assumed linear variation between sample ordinates of the form

$$(1/2)T[v_1 + v_2]$$

where T is the 'width' of the contribution and v_1 and v_2 are the values at the ends of the contribution. Performing this piecewise integration for each of the apertures centred on *A,B,C,D,* etc. in turn, the light outputs at the positions *A,B,C,D,* etc. are proportionally given by

$$L_A = (1/8)Z + (3/4)A + (1/8)B$$

$$L_B = (1/8)A + (3/4)B + (1/8)C$$

$$L_C = (1/8)B + (3/4)C + (1/8)D$$

$$L_D = (1/8)C + (3/4)D + (1/8)E$$

showing that the conversion from discrete samples to a continuous signal and back causes each ideal value to be diluted by a 25% contribution, in the case of linear interpolation, from the average of neighbouring samples. This effect is very much like 'boxcar' distortion in a digital-to-analogue converter and can be compensated by frequency equalisation. Had the shutter angle been less than 360 degrees, the distortion would have been less at the expense of reducing the signal-to-noise ratio.

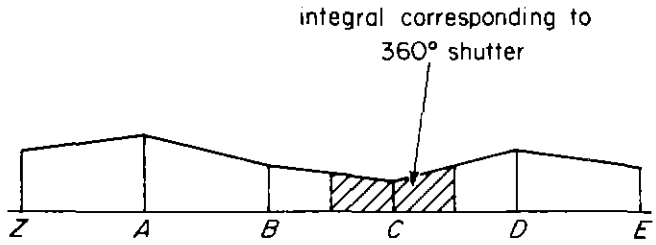


Fig. A1.1 - The signal obtained by linear interpolation of samples from a discrete source.

Clearly there are four combinations of discrete/continuous sources with discrete/continuous displays which will give slightly different results but these variations will be second-order effects. It will be found convenient to develop the arguments in terms of discrete scanning, with appropriate modifications for continuous scanning. If the system is compatible then the quality of the result will be dependent only on the poorer of the source and the display. Thus if it is downwards compatible, varying the source resolution beyond that of the display will make no difference whereas if it is upwards compatible, varying the display resolution beyond that of the source will make no difference. It is sufficient therefore to consider a scanning order disparity between source and display of one.

A1.2 Downwards compatibility

Fig. A1.2 shows a part of the source scan, marked by the points 1,2, etc. together with a part of the display scan shown by the points *A,B,* etc. Fig. A1.3 shows the timing relationship between the source samples

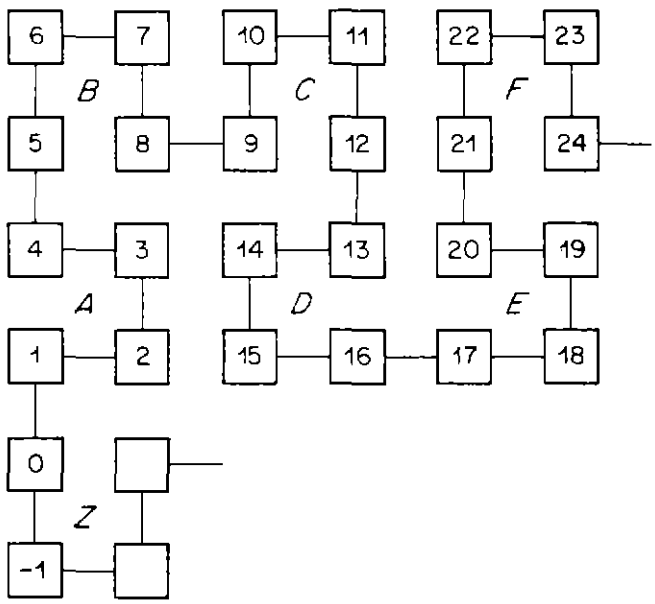


Fig. A1.2 - The spatial relationships involved in upwards and downwards compatibility for a discrete display.

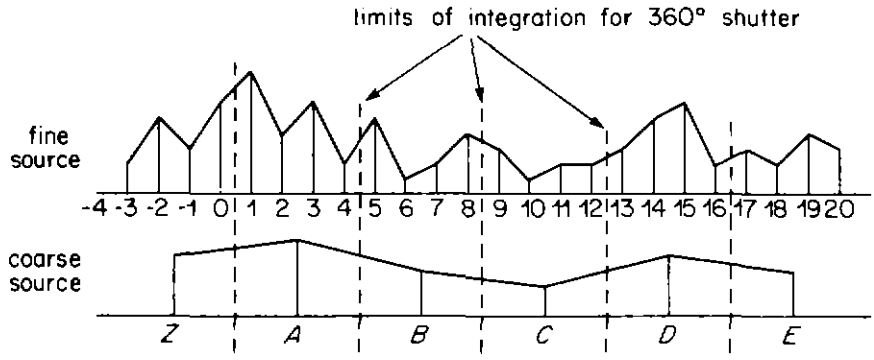


Fig. A1.3 - The timing relationships involved in upwards and downwards compatibility.

1,2,3,4, etc. and those expected from a matched source, *A,B,C,D,* etc. The incoming samples have been linearly interpolated for the purposes of explanation as before. If the system is downwards compatible then the display of 1,2,3,4, etc. should be no different from the display of *A,B,C,D,* etc.

Suppose that the display shutter is 360 degrees, giving an integration time of one coarse pixel period. Performing the integration for each of the apertures centred on *A,B,C,D,* etc. in turn, the light outputs at the positions *A,B,C,D,* etc. are proportionally given by

$$L_A = (1/32)s_0 + (7/32)s_1 + (1/4)s_2 + (1/4)s_3 + (7/32)s_4 + (1/32)s_5$$

$$L_B = (1/32)s_4 + (7/32)s_5 + (1/4)s_6 + (1/4)s_7 + (7/32)s_8 + (1/32)s_9$$

$$L_C = (1/32)s_8 + (7/32)s_9 + (1/4)s_{10} + (1/4)s_{11} + (7/32)s_{12} + (1/32)s_{13}$$

$$L_D = (1/32)s_{12} + (7/32)s_{13} + (1/4)s_{14} + (1/4)s_{15} + (7/32)s_{16} + (1/32)s_{17}$$

where s_n is the value of the nth source sample.

On the other hand, from a matched resolution source, sample *A* would have been the average of samples 1, 2, 3 and 4, sample *B* would have been the average of samples 5, 6, 7, 8 and so on by integration. This shows that, except for a small edge effect, the light output at each point is a superposition of the appropriate values from the fine source and therefore equal to the corresponding matched source values.

Nevertheless, after 360 degree display integration of the matched source signal, the light outputs would have been given by those expressions derived earlier. Comparing the two situations it can be seen that the display of the higher order source actually gives a closer approximation to the matched source samples but this is a second-order effect. Thus downwards compatibility is achieved to the first order.

Fig. A1.4 shows the spatial situation for a continuous display. The locus of the display joins points which could, for the purposes of illustration, be obtained by four-fold averaging of the higher-order source scan points. These points are reached at equal intervals of time and incoming sample values are assumed to be displayed at them. The arrows show the shift in spatial position of the samples brought about by this process, assuming that the

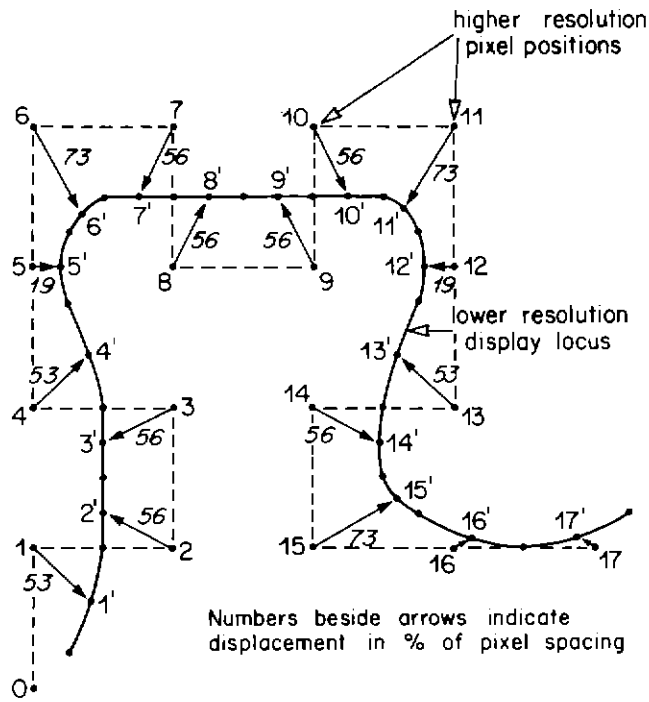


Fig. A1.4 - The mechanism of lower-order display compatibility by superposition for continuous scanning.

scan is optimally synchronised to the video signal. (The figure shows one of the two possible optimum conditions in this case.) It can be seen that the shift never amounts to more than three-quarters of a source pixel and the superposition is approximately achieved by the relatively close juxtaposition of the samples.

A1.3 Upwards compatibility

Now the source order is one lower than the display. Figs. A1.2 and A1.3 will serve to show the relationships between input and display provided the rôles are reversed. Now *A,B,C,D,* etc. represent the incoming samples and 1,2,3,4, etc represent the samples from a matched source. Again assuming a 360 degree shutter, giving an integration time of one fine pixel period, and performing the integration at each of the points 1,2,3,4, etc. on the incoming signal, the light values at the points are as shown in Table A1.1.

Table A1.1

Display point	Output	Display point	Output
1	$(3/8)Z + (5/8)A$	9	$(3/8)B + (5/8)C$
2	$(1/8)Z + (7/8)A$	10	$(1/8)B + (7/8)C$
3	$(7/8)A + (1/8)B$	11	$(7/8)C + (1/8)D$
4	$(5/8)A + (3/8)B$	12	$(5/8)C + (3/8)D$
5	$(3/8)A + (5/8)B$	13	$(3/8)C + (5/8)D$
6	$(1/8)A + (7/8)B$	14	$(1/8)C + (7/8)D$
7	$(7/8)B + (1/8)C$	15	$(7/8)D + (1/8)E$
8	$(5/8)B + (3/8)C$	16	$(5/8)D + (3/8)E$

The average of the four values closest to B is

$$(1/8)A + (3/4)B + (1/8)C$$

whilst the average of the four values closest to C is

$$(1/8)B + (3/4)C + (1/8)D$$

These values are identical to the corresponding values that would have been obtained with a matched display as derived above and it is seen that the effect of the higher-order display is to impose a gradual transition between the values B and C at points in the vicinity of B and C. Bearing in mind the fact that the samples B,C etc. are, themselves, samples of a gradually changing image, this effect is not inappropriate although it may not be rigorously correct. Thus compatibility is achieved.

A continuous display behaves, in this instance, like a discrete display since the signal will be changing relatively slowly, owing to the comparative lack of resolution conferred by the source.

APPENDIX 2

The Mechanism of Hierarchical Compatibility for Interlace

A2.1 Sequential/interlace compatibility

Consider first a discrete interlaced display with a sequential source. Fig. A2.1 shows the timing relationship between the incoming samples 1,2,3,4, etc. and those expected on both fields from an interlaced source wherein samples 1,3,5, etc. would be input on the first field and samples 2,4,6, etc. on the second. As before the incoming samples have been linearly interpolated for the purposes of explanation.

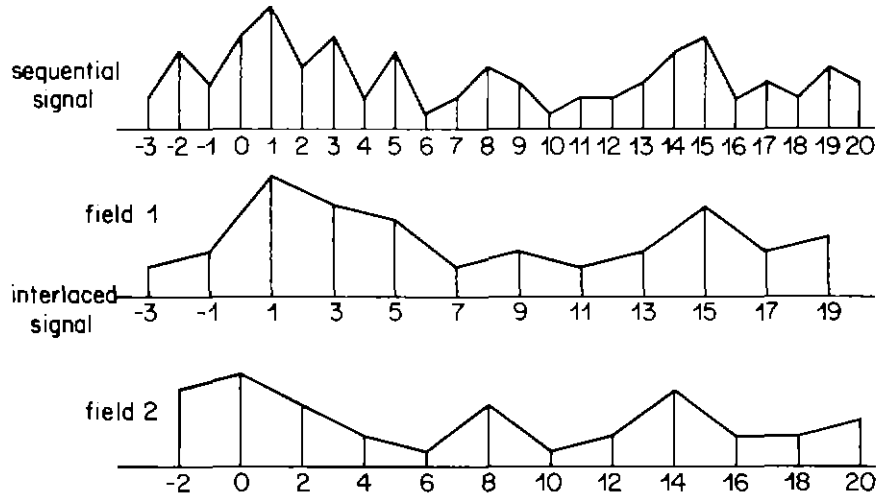


Fig. A2.1 - The timing relationships involved in sequential/interlace compatibility.

Suppose that the display shutter is 360 degrees so that the apertures of adjacent display pixels abut in time, giving an integration time of two sequential pixels. Performing the integration for each of the apertures centred on 1,3, etc. and 2,4, etc. in turn, the light outputs at the positions 1,2,3,4, etc. are proportionally given by

$$L_1 = (1/4)s_0 + (1/2)s_1 + (1/4)s_2$$

$$L_3 = (1/4)s_2 + (1/2)s_3 + (1/4)s_4$$

on the first field and

$$L_2 = (1/4)s_1 + (1/2)s_2 + (1/4)s_3$$

$$L_4 = (1/4)s_3 + (1/2)s_4 + (1/4)s_5$$

on the second field, showing that each correct value is diluted by an equal amount of the average of its neighbours in the other field. This is undesirable as it leads to resolution loss and can be mitigated by reducing the display shutter angle. For example, reducing it to 180 degrees reduces the dilution by 50%.

In comparison, the light outputs for a signal from a matched interlaced source are given by

$$L_1 = (1/8)s_{-1} + (3/4)s_1 + (1/8)s_3$$

$$L_3 = (1/8)s_1 + (3/4)s_3 + (1/8)s_5$$

on the first field and

$$L_2 = (1/8)s_0 + (3/4)s_2 + (1/8)s_4$$

$$L_4 = (1/8)s_2 + (3/4)s_4 + (1/8)s_6$$

on the second field, showing that the interlaced signal also suffers an aperture loss caused by a 3:1 dilution with

the average of the neighbours on the same field. A 180 degree display shutter reduces this to a 7:1 dilution and so is beneficial for interlace sources also.

Consider now a continuous interlaced display. If the display operates according to the first method where it traces the full Hilbert curve and black samples are inserted into the signal, then these effectively obliterate the alternate extra samples generated by the sequential source, making it indistinguishable from an interlaced source. The display then approximates to a discrete display with a 180 degree shutter, the degree of approximation depending on the profile of the black sample insertion and the shape of the beam. If, however, the display operates according to the second method where it traces only the appropriate points and there is no black sample insertion, then integration causes the light at each point to be diluted by neighbouring samples. Now the display approximates to a discrete display with a 360 degree shutter, the degree of approximation depending on the proportion of time spent between the points. It is for this reason that the first method is preferred.

In either discrete or continuous cases it can be seen that compatibility is preserved to a varying degree and, as shuttering is equivalent to black sample insertion, it appears that this technique is applicable to discrete displays as well as continuous displays.

Now consider an interlaced source with a discrete sequential display. Fig. A2.1 will serve to show the timing relationships, reversing the roles of input and display. Now the integration time is one sequential pixel period and all positions are active on both fields. Performing the integration, the light outputs at the points 1,2,3,4, etc. are given by

$$L_1 = (1/16)s_{-1} + (7/8)s_1 + (1/16)s_3$$

$$L_2 = (1/2)s_1 + (1/2)s_3$$

$$L_3 = (1/16)s_1 + (7/8)s_3 + (1/16)s_5$$

$$L_4 = (1/2)s_3 + (1/2)s_5$$

on the first field and

$$L_1 = (1/2)s_0 + (1/2)s_2$$

$$L_2 = (1/16)s_0 + (7/8)s_2 + (1/16)s_4$$

$$L_3 = (1/2)s_2 + (1/2)s_4$$

$$L_4 = (1/16)s_2 + (7/8)s_4 + (1/16)s_6$$

on the second field, showing that the approximately true values at each point are alternated with the average of the neighbours in the other field. On the other hand, an interlaced display would alternate the true value at each point with black, resulting in more twitter. Integrated over the two fields, the output of the sequential display is somewhat like the 360 degree interlaced display with a sequential source but the difference is that the aperture loss cannot be mitigated by shuttering. Thus, compared with an interlaced display, this display exchanges twitter for resolution on interlaced sources. Whether or not this is considered an improvement is an open question.

A continuous sequential display behaves much like a discrete display in these circumstances since no shuttering is involved.

A2.2 Downwards compatibility

First consider discrete scanning. The spatial positions corresponding to the samples are as in the non-interlaced situation of Fig. A1.2. The timing relationships, however, are shown in Fig. A2.2 This shows the relationship on both fields between the incoming samples, 1,2,3,4, etc. and those expected from a source of the same resolution as the display, A,B,C, etc. Samples 1,3,5, etc. or A,C, etc., are input on the first field and samples 2,4,6, etc. or B,D, etc. are input on the second field. As usual, the samples have been linearly interpolated for simplicity.

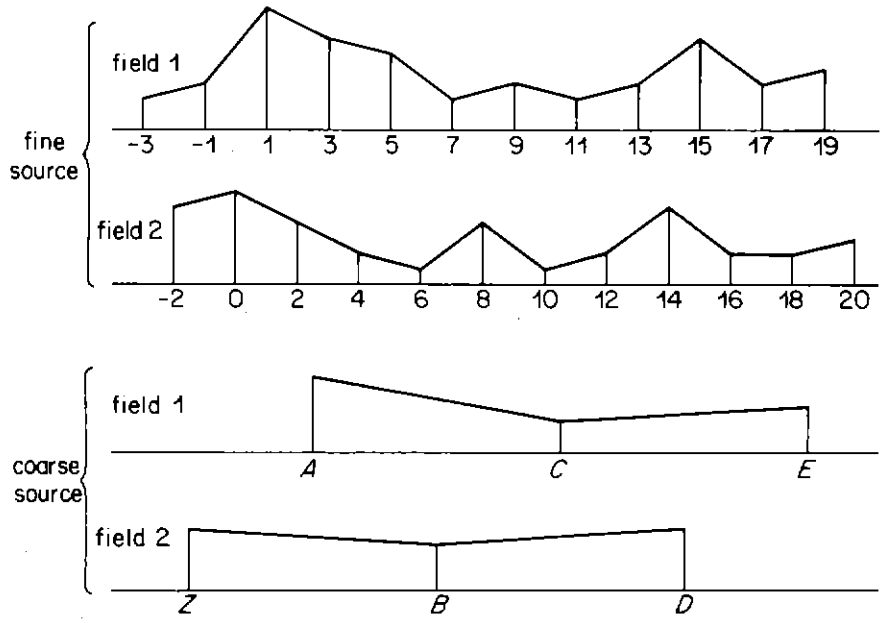


Fig. A2.2 - The timing relationships involved in upwards and downwards compatibility for interlaced scanning.

From the analysis of sequential-interlace compatibility, assume that the display shutter is 180 degrees, giving an integration time of two fine pixel periods. Performing this integration for each of the apertures centred on A,C, etc. and B,D, etc. in turn the light outputs at the positions A,B,C,D, etc. are proportionally given by

$$L_A = (1/64)s_{-1} + (23/64)s_1 + (31/64)s_3 + (9/64)s_5$$

$$L_C = (1/64)s_7 + (23/64)s_9 + (31/64)s_{11} + (9/64)s_{13}$$

on the first field and

$$L_B = (9/64)s_4 + (31/64)s_6 + (23/64)s_8 + (1/64)s_{10}$$

$$L_D = (9/64)s_{12} + (31/64)s_{14} + (23/64)s_{16} + (1/64)s_{18}$$

on the second field. It can be seen that approximately half the light at each position is contributed by one of the four nearest samples with approximately the other half being derived from the two neighbouring samples in the same field, one being dominant. Although the dominant sample is the other nearest to the display position the influence of the third sample introduces an asymmetry in that the centre of gravity of the result is retarded on the first field and advanced on the second. As the shutter is made smaller the result can be made to depend on only the two nearest samples, but one will then have three times the weight of the other and so the asymmetry still remains.

On the other hand, from a matched resolution source, samples *A,B,C,D* would have been input and, after 180 degree display integration would have yielded:

$$L_A = (1/16)Y + (7/8)A + (1/16)C$$

$$L_C = (1/16)A + (7/8)C + (1/16)E$$

on the first field and

$$L_B = (1/16)Z + (7/8)B + (1/16)D$$

$$L_D = (1/16)B + (7/8)D + (1/16)F$$

on the second field, showing that the correct value is only slightly diluted by its neighbours in the same field. As

sample *A* would have been the average of samples 1 to 4, sample *B*, the average of samples 5 to 8 and so on, by aperture integration, the comparison of the display of the two sources is approximately as shown in Table A2.1.

Table A2.1
Sample weightings

	Display output	Sample number							
		1	2	3	4	5	6	7	8
Fine source	*A*	3/8	0	1/2	0	1/8			
	B				1/8	0	1/2	0	3/8
Matched source	*A*	1/4	1/4	1/4	1/4				
	B					1/4	1/4	1/4	1/4

As can be seen the result represents approximately what would have been obtained from the matched source and so, to an extent, compatibility is preserved.

Considering compatibility with higher-order sources, it becomes clear that the optimum shutter angle is 180 degrees. When the source order is much higher than the display the signal, considered on the scale of the display, appears to be sequential rather than interlace and it appears to spend alternating periods sending information about the region of pixels *A,B,C,* etc. in the same field. Clearly, in this situation, the integration must be switched off during alternate periods to avoid information about even pixels diluting information about odd pixels and vice versa.

An important conclusion of this analysis is that, again, with discrete displays, black-level samples must be inserted into the incoming signal at the display pixel rate to preserve the hierarchical property.

Fig. A2.3 shows the corresponding situation for continuous scanning where the samples on one field are

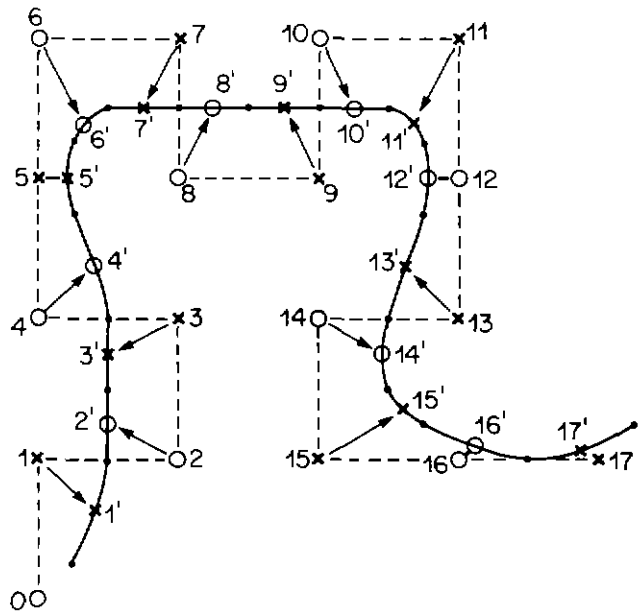


Fig. A2.3 - The mechanism of lower-order display compatibility by superposition for interlaced, continuous scanning.

marked X and those on the other are marked 0. As in Fig. A1.4 the arrows show the shift that is introduced. Again, superposition is achieved by the juxtaposition of samples.

A2.3 Upwards compatibility

Fig. A2.2 will serve to show the timing relationships, reversing the rôles of input and display whilst Fig. A1.2 serves to show the spatial relationships, reversing the roles. Thus samples A,C, etc. and B,D, etc. become the input and samples 1,3, etc. and 2,4, etc. become those expected from a matched source. The 180 degree display shutter width is now half the finer pixel period. Performing the integration for the positions 1,3, etc. and 2,4, etc. in turn, Table A2.2 shows the light outputs as a function of the input samples.

Table A2.2

Field 1		Field 2	
Display point	Output	Display point	Output
1	$(3/16)Y + (13/16)A$	2	$(9/16)Z + (7/16)B$
3	$(15/16)A + (1/16)C$	4	$(5/16)Z + (11/16)B$
5	$(11/16)A + (5/16)C$	6	$(1/16)Z + (15/16)B$
7	$(7/16)A + (9/16)C$	8	$(13/16)B + (3/16)D$
9	$(3/16)A + (13/16)C$	10	$(9/16)B + (7/16)D$
11	$(15/16)C + (1/16)E$	12	$(5/16)B + (11/16)D$
13	$(11/16)C + (5/16)E$	14	$(1/16)B + (15/16)D$
15	$(7/16)C + (9/16)E$	16	$(13/16)D + (3/16)F$

Integrated over both fields, the average of the four samples closest to B is

$$(1/64)Z + (9/32)A + (7/16)B + (7/32)C + (3/64)D$$

whilst the average of the four samples closest to C is

$$(3/64)A + (7/32)B + (7/16)C + (9/32)D + (1/64)E$$

Compared with Table A2.1, it can be seen that the effect of interlace is to dilute the average near a particular source sample with more of the neighbouring values than is the case with non-interlaced scanning.

In addition, however, there are temporal effects because samples 1,3,5,7, etc. are displayed at different times from samples 2,4,6,8, etc. Thus, integrated over single fields, the averages are

	Field 1	Field 2
B region	$(9/16)A + (7/16)C$	$(1/32)Z + (7/8)B + (3/32)D$
C region	$(3/32)A + (7/8)C + (1/32)E$	$(7/16)B + (9/16)D$

showing that the approximately true value is alternated with the approximate mean of the neighbours in the other field. This behaviour is like that of the sequential display with an interlaced source and reduces twitter in exchange for loss of resolution.

On the other hand, with a matched display, the light outputs after 180 degree shuttering would have been

$$L_A = (1/16)Y + (7/8)A + (1/16)C$$

$$L_C = (1/16)A + (7/8)C + (1/16)E$$

on the first field and

$$L_B = (1/16)Z + (7/8)B + (1/16)D$$

$$L_D = (1/16)B + (7/8)D + (1/16)F$$

as derived above. Thus, expressed in the above terms, the light output would be

	Field 1	Field 2
B region	—	$(1/16)Z + (7/8)B + (1/16)D$
C region	$(1/16)A + (7/8)C + (1/16)E$	—

Thus, with the matched display, the brightness alternates between the true value and black, giving rise to twitter and the effect of the higher order display is to reduce the twitter by 6 dB in exchange for loss of resolution. Bearing in mind the marked dependence of 'twitter' on brightness due to the non-linearity of the conventional display, this factor of 6 dB will give a significant improvement.

APPENDIX 3

Derivation of the Scanning Coordinate Sequences

The x and y coordinate sequences for the first-order case of Fig. 5 are shown in Table A3.1, in units of the scan step. The y sequence is a palindrome and the x sequence is a complementary palindrome. The coordinate sequences of the second-order case, written in binary form, are shown in Table A3.2. As there are 4×4 points, each coordinate requires two bits and there are 16 entries. The more significant bit is seen to be a replication of the corresponding first-order pattern, each entry being replicated four times. The less significant bit is seen to be the sequence $YXX\overline{Y}$ for the x coordinate or $XYY\overline{X}$ for the y coordinate, where X and Y are the first-order patterns of the x and y coordinates. The first of these sequences is a complementary palindrome and the second is a palindrome.

Table A3.1

x	y
0	0
0	1
1	1
1	0

Table A3.2

x	y	x	y
00	00	10	10
01	00	10	11
01	01	11	11
00	01	11	10
00	10	11	01
00	11	10	01
01	11	10	00
01	10	11	00

Development of the higher-order cases proceeds along the same rules. The third-order case, shown in Table A3.3, has 64 entries of three-bit values, the sequences of the first two bits of x and y being the same as those

Table A3.3

x	y	x	y	x	y	x	y
000	000	000	100	100	100	111	011
000	001	001	100	101	100	111	010
001	001	001	101	101	101	110	010
001	000	000	101	100	101	110	011
010	000	000	110	100	110	101	011
011	000	000	111	100	111	100	011
011	001	001	111	101	111	100	010
010	001	001	110	101	110	101	010
010	010	010	110	110	110	101	001
011	010	010	111	110	111	100	001
011	011	011	111	111	111	100	000
010	011	011	110	111	110	101	000
001	011	011	101	111	101	110	000
001	010	010	101	110	101	110	001
000	010	010	100	110	100	111	001
000	011	011	100	111	100	111	000

of the corresponding second-order case, each entry being replicated four times. The least significant bit of the x coordinate has the sequence

$$XYY\overline{X} \quad YXX\overline{Y} \quad YXX\overline{Y} \quad \overline{X}\,\overline{Y}\,\overline{Y}X$$

where X and Y represent the same sequences as before. It can be seen that the initial sequence of each group of four follows the same pattern as the sequences in the first group. Likewise, the least significant bit of the y coordinate has the sequence

$$YXX\overline{Y} \quad XYY\overline{X} \quad XYY\overline{X} \quad \overline{Y}\,\overline{X}\,\overline{X}Y$$

It will be noted that the least significant bits are appropriately palindromic or complementary palindromic.

The rule for derivation of the higher-order cases is therefore:

1. The sequences of all but the least significant bit are those of the previous order, each entry being replicated four times.

2. The sequence of the least significant bit consists of groups of four sub-sequences, each taking the form $ABB\overline{A}$.

3. The first sub-sequence of the nth group of four is the same as the nth sub-sequence.

4. A can be X or Y or $\overline{X}$ or $\overline{Y}$. If A is X then B is Y and vice versa.

5. The first sub-sequence of x is X if the order is odd and Y if it is even, and vice versa for y.

6. X and Y are the x and y sequences of the first order case.

APPENDIX 4

Hardware Generation of the Scan

The generation is based on the process indicated in Fig. A4.1. This shows how an $n \times n$ scan can be enlarged to a $2n \times 2n$ scan by reproducing the $n \times n$ scan in sequence in each of the four quadrants of the larger scan. As observed in Section 2, the first and last quadrants are rotated through 90 degrees which can be described as reflection about 45 degree axes as shown. If x and y axes are chosen such that the origin is in the top left-hand corner, as shown, then the reflection of the first quadrant can be obtained by transposing x and y whilst the reflection of the last quadrant can be obtained by transposing x and y and complementing them (subtracting from the highest expressible number). These operations of transposing and complementing are governed by the most significant bits of the x and y coordinates, X and Y, which determine which quadrant the pattern is in. The operations as a function of the quadrant are shown in Table A4.1.

Table A4.1

Quadrant	Y	X	Transpose	Complement
A	0	0	yes	no
B	0	1	no	no
C	1	1	no	no
D	1	0	yes	yes

The operations are seen to be

$$T = \overline{X}$$

and

$$C = Y\overline{X}$$

and are applied to all lower bits of the coordinates. Thus, a building block can be derived as shown in Fig. A4.2 wherein these operations are carried out on the n bits of the incoming signals in adding the $(n+1)$th bits. The building blocks increase in size as more bits are added and Fig. A4.3 shows how they are connected together to form eight-bit outputs.

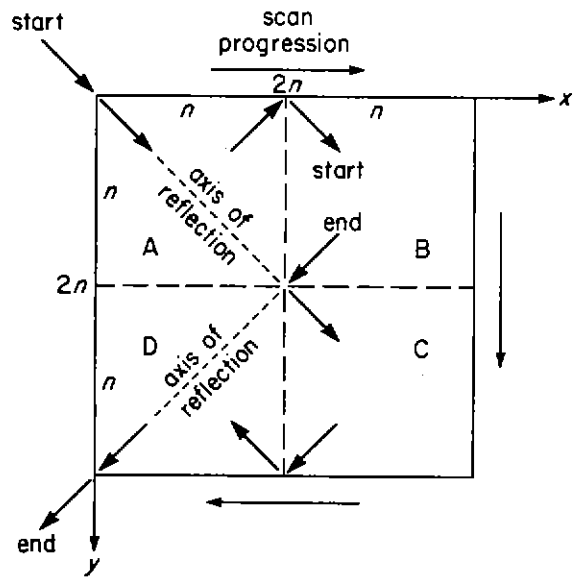


Fig. A4.1 - Development of the 2n-by-2n scan from the n-by-n scan.

The x and y input pairs correspond to the outputs of successive stages of a chain of divide-by-four counters, each operating in the sequence given in Table A4.1 which is, in fact, a gray code sequence, with the 'carry' generated at the end of that sequence. This can be implemented with pairs of normal binary counter stages, the x bit being the EXCLUSIVE-OR of the lower and upper bits of the pair and the y bit being the upper bit alone. This constitutes a two-bit binary-to-gray code converter.

Fig. A4.3 shows that the lower significance bits may suffer many repeated operations of transposition and complementation. As each of these operations is its own inverse an even number of operations has no effect whereas an odd number has the same effect as one alone. So, to simplify the logic, the number of times each operation is required for each particular x, y pair can be totalled, modulo 2, and this resultant control signal can be applied to a single operating unit for that pair. This is shown in Fig. A4.4.

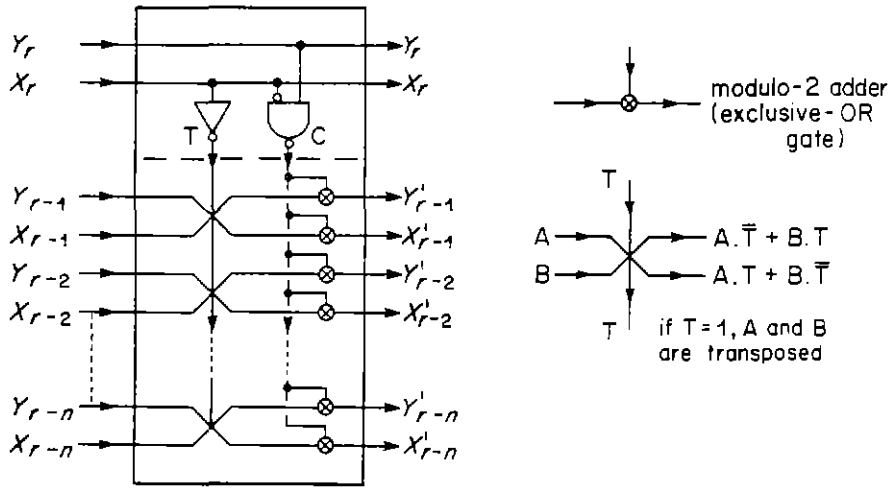


Fig. A4.2 - A variable-sized generator building block.

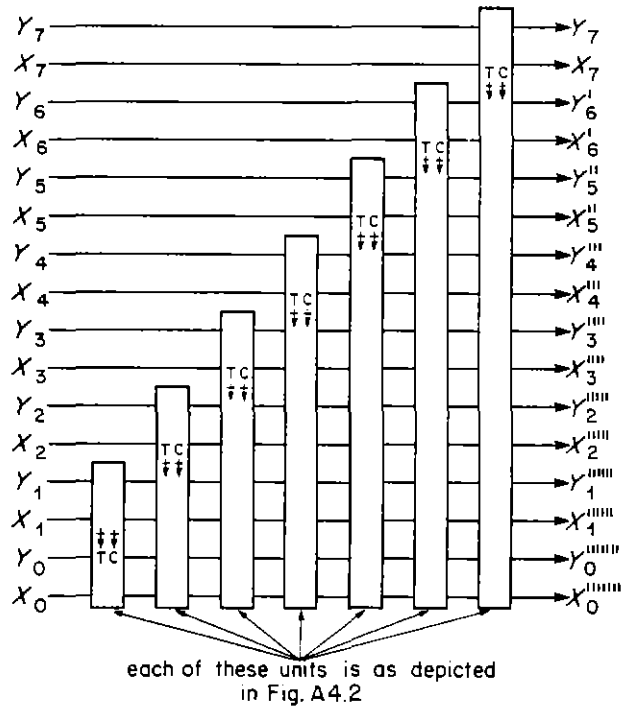


Fig. A4.3 - Cumulative application of the variable-sized building blocks to form eight-bit signals.

Fig. A4.4 also indicates an area of the circuit comprising eight inputs and eight outputs which could be regarded as an alternative building block, this time of fixed size. This function could be implemented as a 256 × 8 ROM programmed with all possible input combinations and the corresponding outputs. It could also include the conversion of the input pairs from binary counter sequence to gray code sequence. Only three such identical ROMs would be needed to produce the complete function for nine-bit x and y inputs. In practice, the generator circuit, although being based on Fig. A4.4, was not implemented with ROMs because of the need for flexibility and the requirement to produce an endless scan.

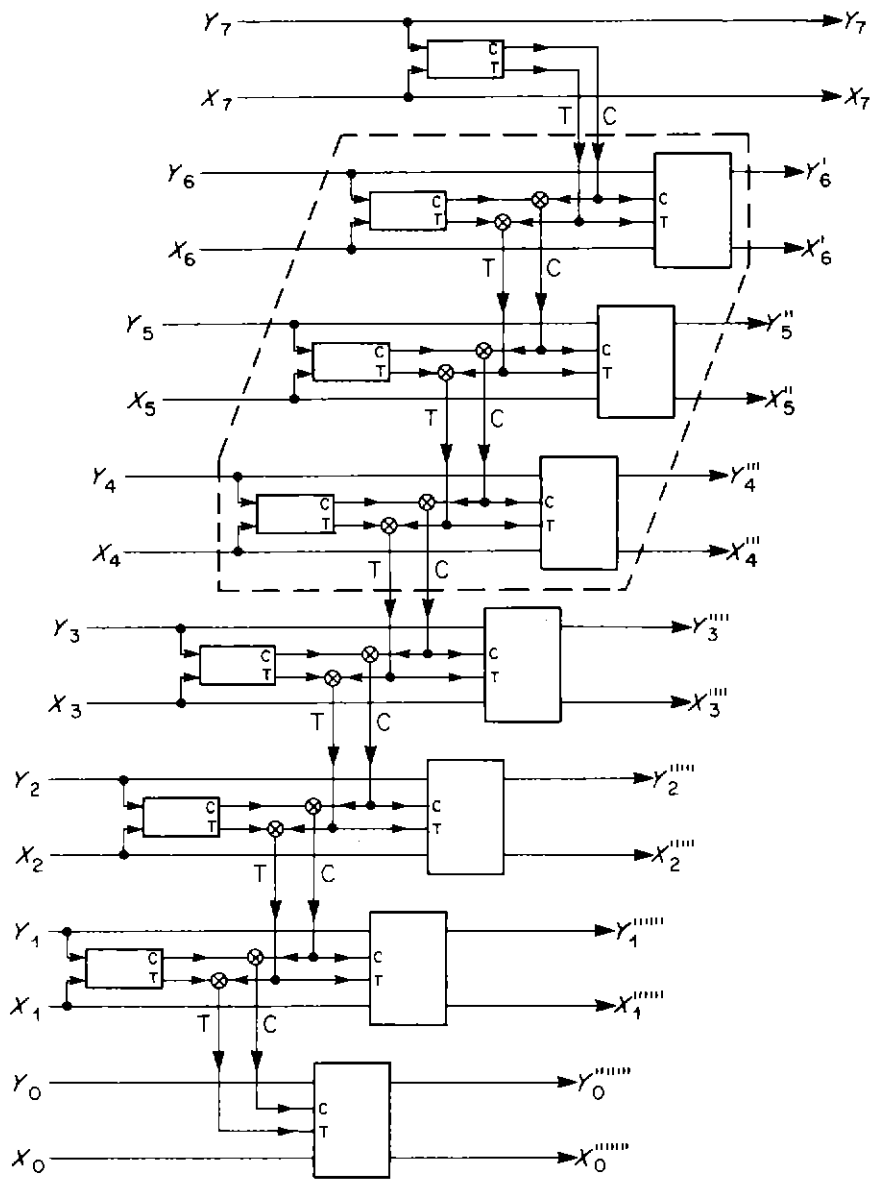


Fig. A4.4 - A simplified arrangement resulting in a fixed size building block.

Printed by BBC RESEARCH DEPARTMENT. Kingswood Warren, Tadworth, Surrey, KT20 6NP